

Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2005, 19:08:10 ; Search time 42 Seconds (without alignments)

1459.288 Million cell updates/sec

Title: US-10-623-429-9

Perfect score: 3355

Sequence: 1 MASAEWRLLEAPLDPRAVP.....VNVDTARAADLFVSSQMMGSR 637

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0‡ Maximum Match 100‡

Listing first 45 summaries

Database : PIR\_79:\*

1: pi1:\*

2: pi2:\*

3: pi3:\*

4: pi4:\*

‡ Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score  | Query | Match | Length | DB ID  | Description          |
|------------|--------|-------|-------|--------|--------|----------------------|
| 1          | 2333   | 68.9  | 635   | 1      | WMBE6  | capsid protein - h   |
| 2          | 1262   | 37.3  | 306   | 2      | WMBE6  | HSV 1 protease -     |
| 3          | 1079.5 | 31.9  | 646   | 1      | WZBEC  | 68.6K capsid protein |
| 4          | 1056   | 31.2  | 647   | 2      | T42579 | capsid protein 35    |
| 5          | 943    | 27.9  | 522   | 2      | SS2216 | viral protease -     |
| 6          | 824.5  | 24.4  | 605   | 1      | WZBES  | gene 33 protein -    |
| 7          | 401    | 11.8  | 586   | 1      | A43675 | capsid protein p40   |
| 8          | 383    | 11.3  | 643   | 2      | SS5610 | polyprotein - equi   |
| 9          | 372.5  | 11.0  | 605   | 1      | QQBES  | BVRP2 (BC-RF1) pro   |
| 10         | 331    | 9.8   | 499   | 2      | T42930 | hypotheical prote    |
| 11         | 322    | 9.5   | 590   | 1      | WMBECB | 64K capsid assemb    |
| 12         | 317.5  | 9.4   | 708   | 1      | QQBESB | UL80 protein - hum   |
| 13         | 301.5  | 8.9   | 475   | 1      | WZBEM4 | gene 17 protein -    |
| 14         | 210    | 8.3   | 839   | 2      | T04859 | extensin homolog F   |
| 15         | 278    | 8.2   | 524   | 2      | T03112 | minor capsid staff   |
| 16         | 272    | 8.2   | 697   | 2      | JC2365 | fused protease-c     |
| 17         | 270.5  | 8.1   | 528   | 2      | T44013 | protease / in-Er     |
| 18         | 270.5  | 8.0   | 528   | 2      | T44200 | probable proteinas   |
| 19         | 269    | 7.9   | 620   | 2      | S06733 | hydroxyproline-ric   |
| 20         | 264    | 7.8   | 2715  | 2      | T13049 | eyelid - fruit fly   |
| 21         | 262.5  | 7.8   | 519   | 2      | S55410 | protease - human     |
| 22         | 258.5  | 7.6   | 760   | 2      | T06291 | extensin homolog T   |
| 23         | 254.5  | 7.5   | 574   | 2      | T43556 | Wiskott-Aldrich sy   |
| 24         | 254.5  | 7.5   | 574   | 2      | T38819 | atrophin-1 - human   |
| 25         | 253.5  | 7.5   | 1184  | 2      | G01763 | extensin - Volvox    |
| 26         | 251    | 7.4   | 464   | 2      | S22697 | protein tyrosine p   |
| 27         | 250    | 7.4   | 1494  | 2      | T1455  | atrophin-1 - human   |
| 28         | 247.5  | 7.3   | 1184  | 2      | S50812 | extensin-like prot   |
| 29         | 246.5  | 7.3   | 1188  | 2      | S49915 |                      |

## ALIGNMENTS

| RESULT 1   |   |   |             |              |             |      |
|--|---|---|-------------|--------------|-------------|------|
| WMBE6  |   |   |             |              |             |      |
| C:Species:   | capsid protein - human herpesvirus 1  |   |             |              |             |      |
| C:Date:  | 31-Dec-1989   | #sequence_revision  | 31-Dec-1989 | #text_change | 09-Jul-2004 |      |
| C:Accession:   | H30084  |   |             |              |             |      |
| R:McGeoch, D.J.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Pzame, M.C.; McNab, D.; Perr |   |   |             |              |             |      |
| J. Gen. Virol. 69, 1531-1574, 1988   |   |   |             |              |             |      |
| A:Title:   | The complete DNA sequence of the long unique region in the genome of herpes sim |   |             |              |             |      |
| A:Reference number:  | A30083; PMID: 88274327;   |   |             |              |             |      |
| A:Status:  | translation not shown   |   |             |              |             |      |
| A:Molecule type:   | DNA   |   |             |              |             |      |
| A:Residues:  | 1-635 < MCG>  |   |             |              |             |      |
| A:Cross-references:  | UNIPROT: P10210; GB:X14112; PID: 91944536; PMID: 2839394                        |   |             |              |             |      |
| C:Genetics:  |   |   |             |              |             |      |
| A:Gene:  | UL26  |   |             |              |             |      |
| C:Keywords:  | varicella-zoster virus gene 33 protein  |   |             |              |             |      |
| Query Match  | 68.9‡   | Score   | 2333        | DB 1;        | Length      | 635; |
| Best Local Similarity  | 74.0‡   | Pred. No.   | 6 6e-114;   |              |             |      |
| Matches  | 477;  | Conservative  | 30;         | Mismatches   | 120;        | Gaps |
| QY   | 1   | MASAEWRLLEAPLDPDRAPVYAGFLALYDSDPGELALDPDTYRAALPPENPLPINVD   | 60          |              |             |      |
| Db   | 1   | MAADAPGDRMEEEPIDPDRAPVYAGFLALYDSDPGELALDPDTYRAALPPDPINVD    | 60          |              |             |      |
| QY   | 61  | HRARCEVGRVLAVYNDPAGPFFVGLIACVQLERVLTEASAAIFERRGPALSREERLLYL | 120         |              |             |      |
| Db   | 61  | HRAGCEVGRVLAVYNDPAGPFFVGLIACVQLERVLTEASAAIFERRGPPLSREERLLYL | 120         |              |             |      |
| QY   | 121   | ITNYLPSVSLSTKRGDDEVPPDSTLFAHVALCAIGRRGTTIVYDTSILDAIAFPRHIDP | 180         |              |             |      |
| QY   | 121   | ITNYLPSVSLATKRGGEAHAPDRTLFAHVALCAIGRRGTTIVYDTSILDAIAFPRHIDP | 180         |              |             |      |
| QY   | 181   | ATPEGVRESEAEEELALGRTWAPGVEAHTLILSTAVNMMLRDVSLVAERQRAGIA     | 240         |              |             |      |
| Db   | 181   | ASREGARLLAAEELAISGRTWAPGVEAHTLILSTAVNMMLRDVSLVAERQRAGIA     | 240         |              |             |      |
| QY   | 241   | GHTYQASEKEFKIWGAEASAPAPERYKTKGAPGAMDTSPAAVSPAPQ---VAVTRQVAS | 296         |              |             |      |
| Db   | 241   | GHTYQASEKEFKIWGAEASAPAPERYKTKGAPGAMDTSPAAVSPAPQ---VAVTRQVAS | 296         |              |             |      |
| QY   | 297   | SSSSSFAPADINPVSASAGAPAPPDGDYIWI PASHYNQLVGAPRHPPLTAGCL      | 356         |              |             |      |
| Db   | 301   | S-----PVLPKMNPPVPTSGTAPAPPFGDGSVWIPASHYNQLVGAPQPHSAFGP      | 354         |              |             |      |
| QY   | 357   | P-AAGTYAIGHPEAGPSHYPVPPPAHPPGMLFAGPSPLEAQAAALVGAIAADRQAGGLP | 415         |              |             |      |
| Db   | 355   | PAAGSVAYGPHAGLSOHPPVHQAQYPPVLFSGPSPLEQIAALVGAIAADRQAGQF     | 414         |              |             |      |

Qy 416 AAGDGHGTRGSAKRRHVEQPPYDCCRDEPDRDFPYPPGEARPEPRVDSRRAARQASG 475  
 Db 415 -ARGDPGTRGSGKRRTEAQPSESYCQDPEADPYPPGEARPGDQRGDSRRAARQSPG 473  
 Qy 476 PHETITALVGAVTSQOELAHMBARTAHPYGYPPYPPGPPYHPHADTE-TPAQPPRYPKA 534  
 Db 474 TNETITALMGAVTSQOELAHMBARTAHPYGYPPYHPHADTE-TPAQPPRYPKA 533  
 Qy 535 VILPPPHIAPPGPPLSGA-VVPPPSYKPPVAVTPGPPAQLHQSPAHHPPPPCTPP 592  
 Db 534 VYRPPHESPAVYGPQGPASHAPVYPAACPGPPP--PCPSSTOTRAPLPTEAFFP 590  
 Qy 593 AASLSPQPEAAGAEAGLVNASSAAHVNVTARAADLVEQVSMQGSR 637  
 Db 591 AATGSPQPEASNAEAGLVNASSAAHVNVTARAADLVEQVSMQGAR 635

Query Match Score 31.9%; Score 1079.5; DB 1; Length 646;  
 Best Local Similarity 40.3%; Pred. No. 5.6e-49;  
 Matches 278; Conservative 75; Mismatches 212; Indels 125; Gaps 20;

Qy 18 AVPIVYAGFLAYDSDPGEELALDPDTYRAALPPENPLPINVDHARCEYGRVLAVVNDP 77  
 Db 12 SLPIVYAGLYDMDGGGEALPPIIDHRRGCVVSBVLSTVDDA 71

Qy 78 RGPFFFLQIACQLERVLETRASAAFERRPALSSEERLLILTVLPSLSSLTRRGD 137  
 Db 72 RGPFFFLQIACQLERVLETRASAAFERRPALSSEERLLILTVLPSLSSLTRRGD 131

Qy 138 EVPPDPDTLFAHVALCAGRRGTTVYVYDTSILDAATAPFRHIDPATEGVREAAEABLAL 197  
 Db 132 DEEPDTLFAHVALCAGRRGTTVYVYDTSILDAATAPFRHIDPATEGVREAAEABLAL 191

Qy 198 A-GRTWAPGVBALHTLSTAVNNNMMLDRNLSVLAARRQAGIAGHTYLOASEKFKCI---  
 Db 192 GDAATWHLSEDTLTRLSTAVNNNMMLDRNLSVLAARRQAGIAGHTYLOASEKFKCI---  
 Qy 254 ---WGAESAPAPRGGYKTOAEGMDSPAASVAPQAVVARQVASSS---  
 Db 252 CNKADFGAELVDT-CGYKSG---SSSFPAADMNPVSASAPAPPGGDOSYLVLPASHNQNLYTQGSARHPP 302  
 Qy 299 ---SSSFPAADMNPVSASAPAPPGGDOSYLVLPASHNQNLYTQGSARHPP 350

Db 303 DKGIGIGERTQHISANASSNPDQTSAGAPLVSGQT-SYILVPAQYQNLVVGQHT-SHPP 359  
 Qy 351 LTACGILPAAGTVYGHGAGGSPHYPPPPAHPYPGMFLFAGP---SPLAQAIALV 402  
 Db 360 INAGPAPVTHA---PSQYIIPAYNSLMPSPSMYQAPWPSVTPHSANLEAQATLV 411

Qy 403 GNAADRGQLPAAQDHGTRGS---AKBRBHEVQPEYDCCGRBEPDR 448  
 Db 412 GALAAORKA---TKGSDPHTVQGSOQSPPLSPQQRERRYARCRHDMDATTED---DL 462

Qy 449 DEPYYPEEARBPBPVDSRRAARQASGHETITALGAVTSLOQETIAHMR-----RT 501  
 Db 463 EGIVYYCQERSRP---GERRAGR---PSITIADLNGAVSSLQQEYISQLRQIQTVAQPO 515

Qy 502 HAPYGPYPPVGPYHPPHADTETPAQPYPYHPPH---IAPPGPPLSGAVVPPP 556  
 Db 516 AAPAGYKPI-----PAPVPOY-SQYQYIQPQHAVAATVAPQLDGI-----P 556

Qy 557 SYPVPAVTPGDPAPLHQSPAHHPPPPFPPTPRAASLQPEPAGAEGALVNASSAA 616  
 Db 557 SQPTQAVLAPQVPGAGPAGPSAKVVAASTAPQAEQARAAPQFEEATSAAVLPVTPQQAS 616

Query Match Score 31.9%; Score 1079.5; DB 1; Length 646;  
 Best Local Similarity 40.3%; Pred. No. 5.6e-49;  
 Matches 278; Conservative 75; Mismatches 212; Indels 125; Gaps 20;

Qy 617 HVNVDTARA-----ADLVEQMGSR 637  
 Db 617 SQTVDSASASTGLEGRDDADIFVQSMSAR 646

RESULT 3  
 W2BEC8  
 68.6K capsid protein - equine herpesvirus 1 (strain Ab4p)  
 C;Species: equine herpesvirus 1  
 A;Note: host: Equus caballus (domestic horse)  
 A;Description: The DNA sequence of equine herpesvirus-1.  
 A;Accession: T36798  
 C;Accession: T36798  
 A;Molecule type: DNA  
 A;Residues: 1-646 <TEL>

RESULT 4  
 T42579  
 capsid protein 35 - equine herpesvirus 4 (strain NS80567)  
 C;Species: equine herpesvirus 4  
 A;Variety: strain NS80567  
 C;Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
 C;Accession: T42579; T42578

A;Cross-references: UNIPROT: P28936; GB: M86664; PIDN: AAB02470-1; PID: 9330827  
 R;Teillard, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.

Virology, 189, 304-316, 1992  
 A;Title: The DNA sequence of equine herpesvirus-1

A;Reference number: A41831; MUID: 929556; PMID: 1318606

A;Contents: annotation; possible protein-coding frames

A;Note: neither amino acid nor nucleotide sequence is given

C;Genetics:

A;Gene: 355

C;Superfamily: varicella-zoster virus gene 33 protein

C;Keywords: capsid protein

|  |    |  |
|--|----|--|
| R.; Telford, B.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.                     | Db | 616 ASSOTVDASASTGLDFGRDDADIFVFSQMSAR 647                                     |
| A; Title: The DNA sequence of equine herpesvirus-4.  |    |  |
| A; Reference number: Z22173; MUID:98264497; PMID:9603135                                       |    | RESULT 5   |
| A; Accession: T42579   |    | 552216   |
| A; Status: Preliminary; translated from GB/EMBL/DBJ  |    | viral proteinase - rabies virus  |
| A; Molecule type: DNA  |    | C; Species: rabies virus   |
| A; Residues: 1-647 <TEL>   |    | C; Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 21-Aug-1998 |
| A; Cross-references: UNIPROT:039278; EMBL:AF030027; NID:92605950; PIDN: AAC059551.1; PID:92    |    | R.; Camacho, A.; Tabaro, B.  |
| A; Accession: T42578   |    | submitted to the EMBL Data Library, June 1994                                |
| A; Status: Preliminary; translated from GB/EMBL/DBJ  |    | C; Accession: 552215   |
| A; Molecule type: DNA  |    | A; Reference number: 552216  |
| A; Residues: 317-647 <TEP>   |    | A; Status: preliminary   |
| A; Cross-references: EMBL:AF030027; NID:92605950; PIDN: AAC059552.1; PID:92605980              |    | A; Molecule type: genomic RNA  |
| A; Genetics:   |    | A; Residues: 1-522 <CAM>   |
| A; Gene: 35  |    | A; Cross-references: EMBL:X79983   |
| A; Function:   |    | A; Note: the source is designated as pseudorabies virus                      |
| C; Superfamily: varicella-zoster virus gene 33 protein   |    | C; Superfamily: varicella-zoster virus gene 33 protein                       |
| C; Keywords: capsid protein  |    | Query Match 27.9%; Score 943; DB 2; Length 522;                              |
| A; Description: N-terminal proteinase domain acts in capsid maturation and is a capsid protein |    | Best Local Similarity 39.5%; Pred. No. 5; 3e-42;                             |
| Matches 261; Conservative 61; Mismatches 155; Indels 184; Gaps 33;                             |    | Matches 261; Conservative 61; Mismatches 155; Indels 184; Gaps 33;           |
| Query 18 AVPIVAGFLALYDSDGPEELDPPDTVALPENPFLINVDHARCEVGRVLAVNDP 77                              | Qy | 20 PIYVAGFLALYDSDCPGEELDPPDTVALPENPFLINVDHARCEVGRVLAVNDP 79                  |
| Best Local Similarity 40.9%; Pred. No. 9; 3e-48;   | Db | 20 PIYVAGFLALYDSDCPGEELDPPDTVALPENPFLINVDHARCEVGRVLAVNDP 79                  |
| Matches 283; Conservative 78; Mismatches 203; Indels 128; Gaps 26;                             | Db | 3 PVVPSGLALYDSDGGEELALTEBIVAAALPPAGLPINIDHARDGAVLAVDDRG 61                   |
| Query 12 SLPIVAGLIALYDNGDGBELTLTRETVAAALPPASRLINIDHNRGCVGEVLSIVDDV 71                          | Qy | 80 PFFVGLIACVQLERVLETAASAAIPEERGPALSREBELLYLITNLYPSVSLSKRGD 138              |
| Db   | Db | 62 PFFVGLIACVQLERVLETAASAAIPEERGPALSREBELLYLITNLYPSVSLSKRGD 118              |
| Query 78 RGPPFVGLIACVQLERVLETAASAAIPEERGPALSREBELLYLITNLYPSVSLSKRGD 137                        | Qy | 139 VPPDRTLFAHVALCAIGRLGTIVYDTSLSLDAIAFPFHLDPATREGVYRREAAEALALA 198          |
| Db   | Db | 119 -ADETLEFAHVALCVIGRVGTIVYDASPERAVGFFRELSSGRSLARAESPD--A 174               |
| Query 72 RGPPFGLITNCPOGLAVLTAAGDFFGEMSEGLSEQEELLYLVSNYLPSASLSRRGIP 131                         | Qy | 199 GRTWAPGVEALTHTLSTAVNNMLRDWSLVAERRRQAGHTYLOASEKPKWGAES 258                |
| Db   | Db | 175 ERWHSMSBPLTALLSTAVNNMLRDWSLVAERRRQAGHTYLOASEKPKWGAES 227                 |
| Query 138 EVPDDRTLFAHVALCAIGRLGTIVYDTSLSLDAIAFPFHLDPATREGVYRREAAEALALA 197                     | Qy | 259 APAPERGYKTTGAGMDTSPAAVPAQVAVRARQVASSSSSSSPAPADMNPSASGAP 318              |
| Db   | Db | 228 GLIPKSG--ASPGA-----RAQ-----CGHGSPP 248                                   |
| Query 132 DEEDEDTLEFAHVSLLCIGRVGTIVYDTPENTVAPFKRLSFSVREBLITIREAQSRL 191                        | Qy | 319 APPPGDSSYLWIPASHINQNLVTGQSAPRHPLTAGLPAAGTVAIGHPGAGSPHYPPP 378            |
| Db   | Db | 249 AERTPGD--YIFVPAQYQNLVNR-----PAP-----PAP 275                              |
| Query 198 AG-RTWAPEWHTLTLSTAVNNMLLRWSLVAERRRQAGHTYLOASEKPKFKI -- 253                           | Qy | 379 PAHPYPGMLFAGPSPLBAQIAALVGATAADROAGGLPAAAGDHGIRGSAKRRHEVEQEP 438          |
| Db   | Db | 276 -----SLSBQLGAIVSA-AMDRERRRSRPEP---RPPARKRKY-----310                      |
| Query 192 GDMATWHVSEDTLTRLSTAVNNMLLRWSLVAERRRQAGHTYLOASASFGITNG 251                            | Qy | 439 YDGRDDEPDRDFPYYPGGEARPEPRPVDSRRAAQSGPHETITALVGAVTSLOQELAHMR 498          |
| Db   | Db | 311 -----DDYAQDNAYYPGEAPP-----PRATSVAV-VSSLQREISHLR 346                      |
| Query 254 -----WGAESAPAPERGYKTG-----APGMDTSPAAASP-----APQAV 289                                | Qy | 499 A--RTHAP- YGPYP-----VG- PYHHPHADTEPAQPPR-----PAKAVLPPP-----540           |
| Db   | Db | 347 AQHVRYPFPYAAPQQLPPGAVGHHPHPHPH-HAAGALYPPMAYQPLHAPPSPVVA 405              |
| Query 252 CNKADFCGAEVLGVT--CGYKSGERVRDYSYSRTLGAFAKFTSGSSALPSNSAGQIGE 309                       | Qy | 541 HIAP- -PGPPLSGAVPP-----PSYP-----PVAVTPGPAPPL-----HQP-SPA 577             |
| Db   | Db | 406 HAVPALPGLPASRSGPVAHVPAPQVVPQVQAVPAAAPPRLQQRHAAFAV 465                    |
| Query 290 RARQVASSSSSSPAPADMNPVSGAAPPGGDSVSLWIPASHINQNLVTGQSAPRHP 349                          | Qy | 578 HAHPP- -PGPPGPTPPPAASLQPQPEAPGAEAGALVNASSAAHYVNVTARAADI.FVSNMNGS 636     |
| Db   | Db | 466 QAAAPRADAQPPQVQAVSASR- -PTESPAPIDASSAA-VACORGADIFVSNMGSQ 521             |
| Query 310 RAKHPGMASS-----NPQTLISAAGAPLYSGD- -YILVPAQYQNLVQVQHT- SHP 357                        | Qy | 637 R 637  |
| Db   | Db | 522 R 522  |
| Query 350 PLTACGLPAGTVA-AYGHPGAG---PSHYPPPPPAHPPGMP-LEAQIAALVG 403                             | Qy |  |
| Db   | Db |  |
| Query 358 PINAQQAISHTAHPQYQDAPYNSMVTQYAPPWSP-----PSTNLQAQITAIVG 411                            | Qy |  |
| Db   | Db |  |
| Query 404 AIAADROAGGLPAAAGDHGIRGS-----AKRRRHEVEQPEYDGRDDEPRD 449                               | Qy |  |
| Db   | Db |  |
| Query 412 ALAADRKA-----KSSDSQAVQNSCQSPPLPSQOERRYTRKRHDWDANTRD-----DIE 462                      | Qy |  |
| Db   | Db |  |
| Query 450 FPPYGEAEPPEPRVDSRAARQSGPHETITALVGAVTSLOELAHMR-----RTH 502                            | Qy |  |
| Db   | Db |  |
| Query 463 GIYYPGER--SPREVE- KFAGRST-----TIADLMGAVSLQEVSLQRAIQTVTAQQV 515                       | Qy |  |
| Db   | Db |  |
| Query 503 APGYPYPPGPPYHPPHADTETPAQPPRPEAKYLLPPHI- -APPGPPLISG- -AVPPPSY 558                    | Qy |  |
| Db   | Db |  |
| Query 516 APAGLYKPP-----GVPQOQSYQO-----VHPOHAVAISLISPOSGTQAOQPSV 563                           | Qy |  |
| Db   | Db |  |
| Query 559 -----PPVATPGPAPLQHOPSPAHAPPPPQGPTPPDAASLQPPEAPGAEAGALYNAS 614                        | Qy |  |
| Db   | Db |  |
| Query 615 AAHVNVDTARA-----ADLVEFSQMSGR 637   | Qy |  |
| Db   | Db |  |

RESULT 6

WZBE33 gene 33 protein - human herpesvirus 3

C;Species: human herpesvirus 3, varicella-zoster virus

C;Accession: G221214

R;Davison, A.J.; Scott, J.B.

J;Gen. Virol. 67, 1759-1816, 1986

A;Title: The complete DNA sequence of varicella-zoster virus.

A;Reference number: A27345; PMID:3018124

A;Molecule type: DNA

A;Cross-references: UNIPROT:P09286; EMBL:X04370; NID:959989; PIDN:CAA27916.1; PID:960022

C;Genetics:

C;Superfamily: varicella-zoster virus gene 33 protein

Query Match Score 824.5; DB 1; Length 605;

Best Local Similarity 33.8%; Pred. No. 8.5e-36;

Matches 233; Conservative 92; Mismatches 227; Indels 137; Gaps 28;

Db 1 MAAEADENCEA-----LVGAGIALLYSKDEGELNITPEIVSALPPSKIPNID 51

Qy 1 NASAEMRERLEAPLDPDRAPVYVAGFLALYDSDGDPGELADLPDTVALPPENPLPINVD 60

Db 61 HARCEFCRVLAUNDRGRPFVGLACVQLERVELTANSAALIPEERGPALSREERLL 120

Qy 52 HRKDCVUGVEVTAIEDIRGPFLGLIVRCVQLHAVLPEAAHNFNGNRDTSVLSPLERALYL 111

Qy 121 ITNYLPSVSLSPKRGDDEVPPRTLPAHVALCAIGRLGLTVYDPSLDAAJAPFHLD 180

Db 112 ITNYLPSVSLSPKRLSPNEITPDGNFETHVVALCUGVRGVTVNYDPESSIEPPVLS 171

Qy 181 ATREGVREAAEELALAGRTWAPGVETATHTLLSTAVNNNMRLDRWSLVAERRRAGIA 240

Db 172 ESK-ARLLSLVKDYGILNKWKVUSDEKLAKVLLSTAVNNNMRLDRWDVAKRREAGIM 229

Qy 241 GHTYLQASEKE-----KINGAESAPAPERG-----YKTAGP-----GMDTSDAASVP 283

Db 230 GHVYQLQSTGIGYLARITNVNGYES-KLPNAGVINTATHPGVTPDLSVWIPASHYHQ 288

Qy 284 APOVAVARQVASSSSSSSSPPADINPPVASSG--APAPP--PGDGSYIWIWIPASHYQ 338

Db 289 VPHLKV-TQLCRNDS-----DNAVAGNANSNTSPQPSGSVPGNGFVLIPTAYYSQ 338

Qy 339 LVTGQSAPRHPLTACGLPAAGTVAIYCHPAGCSPSPHYPP-----PPAH-----PY 386

Db 339 LITGQT--RNQDVSTGAPNNQYIYGPYGSHPHAPPNTGGYCPGHGGPXFGPQ- 394

Qy 387 LFAGPSPLEAQIAALVGAIAADRQGLPLAAA-----GHGIGRSKAKRHRHEVEPEYD 440

Db 395 -YPPPNRLEMOMNSAFNAAER--GIDLOTPCVNFPDKTDVRRGKRFKSMQRELD 450

Qy 441 --CGRDBPDRDEP--YYGE-----ARPPPRPVDSRARAQAGPHETITALVGAFT 488

Db 451 SFYSGSGEQMDGEFSPNTYFPEPPTYITHRBBRVSFSYWRHRVSNQHQH--ELAGVVA 507

Qy 489 SLQOELAHMRKTHAPGYPVPPGVYHPPADTETPQOPPRYPAKAVYLPPHIAPPGP 548

Db 508 KLUQEVTELKSON---GTMQPLS--HH---TNIP-EGTGDPRISLKLQLQ-----S 550

Qy 549 LSGAVPPPSXPVAVTPGAPPLHQPSPAHAAHPPPPGTPPPASLIPQEARGAAGA 608

Db 551 VSGLQSSQN-----TTSTPHTDVTGQDVNA 575

Qy 609 LYNASSAAHNVNDTARAADLFYSQMCGR 637

Db 576 VEASSKAPLQGOSTADAMFQANQMYGR 604

Qy 479 TITALVGAVTSLSQELAHMRARTHA 503

Db 510 GINDLQGLITLQKEYAGLJSASNA 534

RESULT 7

A43675

capsid protein p40 - infectious laryngotracheitis virus

C;Species: infectious laryngotracheitis virus

C;Accession: S13444; A43675

R;Griffin, A.M.

A;Title: The complete sequence of the capsid p40 gene from infectious laryngotracheitis virus

A;Reference number: S13444; MUID:90301599; PMID:2163526

A;Accession: S13444

A;Status: preliminary

A;Residues: 1-386 <NUC>

A;Cross-references: UNIPROT:P43984; GB:D00565; NID:9221899; PIDN:BAA00439.1; PMID:9221900

R;Griffin, A.M.; Boursnell, M.B.G.

J;Gen. Virol. 71, 941-950, 1990

A;Title: Analysis of the nucleotide sequence of DNA from the region of the thymidine kinase gene families.

A;Reference number: A43675; MUID:90218031; PMID:2157797

A;Accession: A33675

A;Molecule type: DNA

A;Residues: 1-516 / EP' <CRI>

A;Cross-references: GB:D00565

A;Experimental source: strain Thorne

C;Superfamily: varicella-zoster virus gene 33 protein

C;Keywords: capsid protein

Query Match Score 11.9%; Score 401; DB 1; Length 586;

Best Local Similarity 24.6%; Pred. No. 7.1e-14;

Matches 139; Conservative 86; Mismatches 218; Indels 122; Gaps 20;

Qy 21 IYVAGGITALYDSDGDPG--ELALDPDPVRAALPP-ENPLPINVDHARACEVGRVLAVVNDP 77

Db 10 RGLFCIGUVSTALAPPFLSYVQDDAIFANAEEMVITETEKFLYLSNIPSLSSRL 129

Qy 78 RGPFFYGLIACVQLERVLETAASAAIF--ERRGPALSGREERLYLITNTYLPSPVSLSTKRR 135

Db 70 RGLFCIGUVSTALAPPFLSYVQDDAIFANAEEMVITETEKFLYLSNIPSLSSRL 129

Qy 136 GDEVPDRTLFAHVAIICAIGRPLGTVYDTSLSAIAJAPFHLDPATRE----GYRREA 190

Db 130 EKNEVGKDFPAHVALCELGERGREGTAYIYGTASEAIGAFDLSAIAKEQLYEIAATREKC 189

Qy 191 AEABLLAAGRTWAPGVEALTHLILLSTAVNNNMRLDRWSLVAERRRAGIAGHTYLGASEK 250

Db 190 AEVPRELSP-----PEITPVLMKRCFTHGFLMIGTCLTRREMAVNPKYLQDNE 242

Qy 251 FKIWGAES-SAPAPERYGKTYKTPGAMDTSPASV-----PAFOVAVRQAVASSSSSSFP 304

Db 243 ITIGKIGHSSETPENNAKDRSV-STQATPSDISESQQPSGQTHYFAME-SATCCQFL 299

Qy 305 APADMPVSAASGAPRPPGDSYIWLIPASHYHQ-----

Db 300 QTKN-----GAPSASREDMTYVPPFERYASLIAASARRDNDRPPSPSREFRRSR 350

Qy 341 --TGQCAP-----RHPPLTAGLPAAG----TVAGHP--GAGSPPHYPBP--- 378

Db 351 DSTHRESPGDRIDWPRGFERIPPLSMEPMGPMNHTYRPAIYEDPNFCGRFPYIYQSPAST 410

Qy 379 -PAHP-----YGMMLPAGPSPLE-----AQIAALVGAIAADRQAGGLPAA 418

Db 411 YPVHPPNYYSSNFGQFG--AGTYPQYPSLHEQTVSRDLALISLE----- 455

Qy 419 GDHGIGRSAKBRHRHEVQPEYDGRFDPDFFYYGEARPEPRVDSRRARQASGPHE 478

Db 456 -----KQNKRKSESYSSNNPRERSARTISEND-PYFFGEMVBAKTTIEQQLCEKGPVGS 509

Qy 479 TITALVGAVTSLSQELAHMRARTHA 503

Db 510 GINDLQGLITLQKEYAGLJSASNA 534

|  |  |
|--|--|
| polyprotein - equine herpesvirus 2<br>N; Contains: capsid scaffold protein; proteinase/capsid protein<br>C; Species: equine herpesvirus 2  | Db 538 ASGGVP-----CAPGPGVGEPPQAPQVTVTHNQHQAAPQGGATGATAANVEQRQP 589       |
| C; Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004   | Qy 602 PGAEA 606   |
| C; Accession: S55610; S55611; R; Relford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.   | Db 590 EGGEA 594   |
| J. Mol. Biol. 249, 520-528, 1995   |  |
| A; Title: The DNA sequence of equine herpesvirus 2.<br>A; Reference number: S55694; MUID:9530501; PMID:7783207   |  |
| A; Status: nucleic acid sequence not shown; translation not shown  |  |
| A; Residues: 1-643 <TEL>   |  |
| A; Cross-references: UNIPROT:P52369; GB:U20824; NID:9695172; PID:ARC13804.1; PID:9695188   |  |
| A; Experimental source: strain 86/67   |  |
| A; Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995   |  |
| A; Accession: S55611   |  |
| A; Status: nucleic acid sequence not shown; translation not shown  |  |
| A; Molecule: DNA<br>A; Residues: 308-643 <TEL>   |  |
| A; Cross-references: GB:U20824; NID:9695172; PID:ARC13803.1; PID:9695189   |  |
| A; Experimental source: strain 86/67   |  |
| A; Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995   |  |
| C; Superfamily: varicella-zoster virus gene 33 protein<br>C; P-307; Product: proteinase/capsid protein #status predicted <MAT1><br>F; 308-643/Product : capsid scaffold protein #status predicted <MAT2> |  |
| Score 383.5%; DB 2; Length 643;<br>Best Local Similarity 24.7%; Pred. No. 6.3e-13;   |  |
| Matches 164; Conservative 86; Mismatches 248; Indels 167; Gaps 28;   |  |
| Query Match 13 PLPDR-----VPIVYAGFLALYD-SCDPGELALDDPTVRAALPPENPLPINVDRAR 64   | Query Match 11.0%; Score 372.5%; DB 1; Length 605;                       |
| Db 26 PVPEAAAARSASRASPVYLGFDVFSYPRDSRALYLNPDVG AHLPLPGPIPANEHLQE 85  | Best Local Similarity 24.8%; Pred. No. 2.2e-12;                          |
| Qy 65 CEVGRVLAVVNDPRGPFFVGLIACVQLERVLE-TAASAIFFERGPALSREBFLYLITN 123   | Matches 166; Conservative 79; Mismatches 237; Indels 187; Gaps 28;       |
| Db 86 AHGWTLGLHTRYGLFCVAVITAEEFFDLRCAASSVARTADHHLPPNPTLEMHT 145  |  |
| Qy 124 YLPVPSLSLKRRGDBVP----PORTLPAHVVALCAIGRLGTIVTYDSLDAATAPPRHLD 179   | 17 RAVPIVYAGFLALYDSDGPGE-LALDDPTVRAALPPENPLPINVDRARCEVGRVLAV 74          |
| Db 146 WLPEPLSLS-THPPALPGKGGDTPIFOVALCAMGQRGTAVYGBSLWLSKFTSL 204   | 3 QADSVYVGFFVERPDA-PPKDACLLHDLDLTVSQLPLKPLPLTVEHLPDAPGSVFGLY 61          |
| Qy 180 PATREGVRRREAAELAELAAGRITWAPGVRL-----THTLSTAVVNMMRLDRWSLVA 231   | 75 NDRGPFPFVGLIACVQLERVLETA-ASAAIFERGPALSREBFLYLITNLYDPSVLSTK 133        |
| Db 205 PEERGAIAEGYA-----SPAPSRLPEPHTCSNEILMAKAIDGFKNRLTLK 253  | 62 QSRGLFSSAATSGDFLISLDSIYHCDIAQSQRPLPREFKPLAHLA-WLPSLSSLAS- 119         |
| Qy 232 ERRRQAGIAGHTYLQASEKFKTGAEASAPAPERGYKGTGAPGAMDTSPASVAPQVAVRA 291   | 134 RRGEDEVPPD-----RTLPFAHVALCAIGRLGTIVTYDSLDAATAPPRHLDPATR 183          |
| Db 254 TDKGVAEYKAPTYLKASVQ-----GLPANLDEVDSA-----RG 286   | Db 120 -----LHDIDIPQTADGKLSSFDHVSICALGRRGTAVYGTDLAWLKHFDLDEPLSA 174      |
| Qy 292 ROVASSSSSSSFPAPA----DNKPNPVSASCAPAAPPGGDSYLMIPASHY-----336  | Qy 184 EGYRREAAEELAAGRITWAPGVRLHT-LLSTAVVNMMRLDRWSLVAERRSQAGIAGH 242     |
| Db 287 GEDPPTAAJATPPTADATMMQQQQPAQA-PAGGDBLISVPRSTFMTLQTNLDTM 345  | Db 175 AQLENDAAAK----RESGCPEDPHPLPLTKLAKAIDAFGLRNRVETLRQDRSVAIPIAE 230   |
| Qy 337 NOLVIGQ-----SAPRHPPLTACGLPAAGTVAYGHPGAGPSRHPPPAH-----381  | Qy 243 TYLQASEKFKINGAEASAPAPERGYKGTGAPGAMDTSPASVAPQVAVARQVASSSSSS 302    |
| Db 346 RQTSLGQRFQDIDAPAPQALRVRPPPAFPFV-HPQXYPAFQYHPOQDAQQYLPTVPL 404   | Db 231 SYLKASD-----APDLQKPDKALQSPPPASTD-----PATMLSGNAGEGATAGGSA 277      |
| Qy 382 PYPPMLPQGSPSLEAQIAALVGAIAADRQAGGLPAAAGD--HGIRGSKARRRHEVEQP 438  | Qy 303 -----FPAPADMNPVSASGAPAPPDPGDSYLM 331                              |
| Db 405 PPGGAMPFAPPPLPDFY-----KYGGIPARYGSPVAHPAR-PGKRKR-----445   | Db 278 AACQDLISVPRNTFTMLLQTNLNDQKPRQTPLDYAAPDPPFSHQIAATAPSPYQPGAGAVA 337 |
| Qy 439 YDCGRDEPRDF--PYYPEARPEPRPVDSRRAORQSGPHEITITALYGAITSLQQLELAH 496   | Qy 332 PASHYNQLVQGOSAPRHPPLTACGLPAAGTVAYGHPGA-----GSPSHYPPPPAHP 382      |
| Db 446 -DC-----DBEEFEGPLFPEBI-----HRDVQSLSKSIALQSELKD 480  | Db 338 PAGGY-----FTSPGGTYAGP-ACGDPGAFLAMDAHTYTHPHPH-PPPAYFG 381          |
| Qy 497 MRARTHAPYGPYPPVGPYTHPHADTETPAQPPRPPA-KAVYLP-----HIAAPPGP-P 548  | Qy 383 YPMLFAGSPSPLAQIAALVGAIAADRQAGGLPAAAGDHGIGRSKARRRHEVEQPYDQG 442    |
| Db 481 IK---NSQQFPQPLPQPOQPOQPOQPOQPAFPQAFYTHPAAGDQGYYVRYLNPFQ 537   | Db 382 LPG-LFGPPPFPYGSHL-----RADYYPAPS-----RSNKRKRDPEDEEGG 426           |
| Qy 549 LSGAVPPPBPYPPVAVTPGP-----APBLHQPSPAHAPPDPGTPPPASLPOPEA 601  | Qy 443 RDEPDRDPPYPPGEARPEPRPVDSRRAORQSGPHEITITALYGAITSLQQLELAHMART- 501  |
| Qy 549 LSGAVPPPBPYPPVAVTPGP-----APBLHQPSPAHAPPDPGTPPPASLPOPEA 601  | Db 427 -----LFPGE-----DATYRQDAGLSKSVNETQHTLQLAARRTEL 462                 |
| Qy 549 LSGAVPPPBPYPPVAVTPGP-----APBLHQPSPAHAPPDPGTPPPASLPOPEA 601  | Db 502 ---HAPYCPYPPVGP-YTH-----PHADTETPAQPPRPPAKAVYLPAPPHIAPPGPPLS 550   |
| Qy 549 LSGAVPPPBPYPPVAVTPGP-----APBLHQPSPAHAPPDPGTPPPASLPOPEA 601  | Db 463 SYGHTGVGycPQOGPCTHSGPYGPQHQSVEVPR-----YVYHP-----502               |

|  |  |    |  |
|--|--|----|--|
| Qy   | 551 GAVPPPSYPVAVTPGPAPPLHQSPSPAHAPPPPPGPTPPPAASLHQPEPAGAEGALV 610      | Qy | 577 -AHAHPPPPPPGPTPPPAASLHQPEPAGAEGALVMASSAAHVNVD----TARAADLFV 630     |
| Db   | 503 --PPP-----PTSHQAAQRQ----PPPGTQAPEAHCVAESSTIP--EAGAAG 541           | Db | 436 APHVQQPMQOPPHVQQPRV-LPSTDWPSNTA-QKPSASESVHVEASFVQDPEVQCKLFC 493    |
| Qy   | 611 NASSAAHVN 619  | Qy | 631 SQMM 634   |
| Db   | 542 NSGPREDITN 550   | Db | 494 DELL 497   |
| <b>RESULT 11</b>   |  |    |  |
| WMBECB   |  |    |  |
| 64K capsid assembly protein - simian cytomegalovirus (strain Colburn)                        |  |    |  |
| C;Species: simian cytomegalovirus  |  |    |  |
| C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004                  |  |    |  |
| C;Accession: AA0414; A30176  |  |    |  |
| R;Welch, A.R.; McNally, L.M.; Gibson, W.   |  |    |  |
| J;Virol. 65, 4091-4100, 1991   |  |    |  |
| A;Title: Cytomegalovirus assembly protein nested gene family: four 3'-coterminal transcripts |  |    |  |
| A;Reference number: A40414; MUID:91303658; PMID:1649317                                      |  |    |  |
| A;Description: Primary structure of the herpesvirus attales genome.                          |  |    |  |
| A;Accession: A40414  |  |    |  |
| A;Molecule type: mRNA  |  |    |  |
| A;Cross-references: UNIPROT:PI16046; GB: M64627  |  |    |  |
| A;Status: preliminary; translated from GB/EMBL/DBBJ  |  |    |  |
| A;Molecule type: DNA   |  |    |  |
| A;Residues: 1-499 <A/B>  |  |    |  |
| A;Cross-references: UNIPROT:Q9YTP9; EMBL:AF083424; PIDN: AAC95541.1                          |  |    |  |
| A;Experimental source: strain 73   |  |    |  |
| C;Superfamily: varicella-zoster virus gene 33 protein  |  |    |  |
| C;Accession: A30176  |  |    |  |
| A;Molecule type: mRNA  |  |    |  |
| A;Cross-references: 281-590 <ROB>  |  |    |  |
| C;Superfamily: cytomegalovirus capsid assembly protein                                       |  |    |  |
| C;Keywords: capsid assembly  |  |    |  |
| F;166-590;Product: 46K capsid assembly protein #status predicted <CAC>                       |  |    |  |
| F;281-590;Product: 34K capsid assembly protein #status predicted <CAE>                       |  |    |  |
| F;349-590;Product: 27K capsid assembly protein #status predicted <CAE>                       |  |    |  |
| Query Match 9.5%; Score 322; DB 1; Length 590;   |  |    |  |
| Matches 159; Conservative 75; Mismatches 239; Indels 162; Gaps 30;                           |  |    |  |
| Qy   | 18 AVPIYVAGFLAYLSDGPGEALD-PTDV--RAALDPENPLPINTVDRGCEVGRYLAW 74         | Qy | 18 AVPIYVAGFLAYLSDGPGEALD-PTDV--RAALDPENPLPINTVDRGCEVGRYLAW 74         |
| Db   | 2 ADPVTGGFLRYDE PPGEAELFLSGVYDWRLRCRGPPLVNHYDSEATVGYAGLQ 60            | Db | 2 ADPVTGGFLRYDE PPGEAELFLSGVYDWRLRCRGPPLVNHYDSEATVGYAGLQ 60            |
| Qy   | 75 NDPRGPFVFFGVLJACYQLERVLETTAAASAAIPERRGP--ALSREERLLYLITNLYLPSVSL 131 | Qy | 75 NDPRGPFVFFGVLJACYQLERVLETTAAASAAIPERRGP--ALSREERLLYLITNLYLPSVSL 131 |
| Db   | 61 NVRAGLFLGRVTSPFLDVTQKASSKSGEVSQGRRTAVLPSVSL 120                     | Db | 61 NVRAGLFLGRVTSPFLDVTQKASSKSGEVSQGRRTAVLPSVSL 120                     |
| Qy   | 139 VPPDRTLFAHVVALCAIGRRLGTVYDTSTLDAIAAPFRHLDPATREGVRREAAELALA 198     | Qy | 139 VPPDRTLFAHVVALCAIGRRLGTVYDTSTLDAIAAPFRHLDPATREGVRREAAELALA 198     |
| Db   | 123 NEENDNNFQHVSLCALGRRGRTVAVYVSMNLKDIAISQFYSISQEVSYDQSKNIDLNL 182     | Db | 123 NEENDNNFQHVSLCALGRRGRTVAVYVSMNLKDIAISQFYSISQEVSYDQSKNIDLNL 182     |
| Qy   | 199 GRTWAPGVELTHLTLSTAVNNMMLRDRWSLVAERRQAGLHTYLOASEKFKingAES 258       | Qy | 199 GRTWAPGVELTHLTLSTAVNNMMLRDRWSLVAERRQAGLHTYLOASEKFKingAES 258       |
| Db   | 183 SK---PVNNMNPYTLMAKIDAGFTKDLRQLRTDGVAQIKKLTVKASETPKL---- 234        | Db | 183 SK---PVNNMNPYTLMAKIDAGFTKDLRQLRTDGVAQIKKLTVKASETPKL---- 234        |
| Qy   | 259 APAPERGKTKGFGAMDTSPAAVSPAPQAVARQYQASSSSSSFPAPDMNPVSASGP 318        | Qy | 259 APAPERGKTKGFGAMDTSPAAVSPAPQAVARQYQASSSSSSFPAPDMNPVSASGP 318        |
| Db   | 235 -----GIEDS-----EDNQHGTIS-- 250                                     | Db | 235 -----GIEDS-----EDNQHGTIS-- 250                                     |
| Qy   | 319 APPPGDGSSYLNPASHY----NQLYTQGQSPAPR---HPLPTAGCLPAAAGTVA---YGH 366   | Qy | 319 APPPGDGSSYLNPASHY----NQLYTQGQSPAPR---HPLPTAGCLPAAAGTVA---YGH 366   |
| Db   | 251 ----QGSDDLISVPKSTPLSMQLNNLNFQVTRPVFYPY---LPSQGTYYVPELYNH 302       | Db | 251 ----QGSDDLISVPKSTPLSMQLNNLNFQVTRPVFYPY---LPSQGTYYVPELYNH 302       |
| Qy   | 422 GIRGSAKRHRVEQPEYDGRDDEPDRDFPYGEARPEPRPVDSRRAARQASGPHEIT 481        | Qy | 422 GIRGSAKRHRVEQPEYDGRDDEPDRDFPYGEARPEPRPVDSRRAARQASGPHEIT 481        |
| Db   | 328 ---PNKREDT---DC-----VFPGE-----SSLXKDVL 352                         | Db | 328 ---PNKREDT---DC-----VFPGE-----SSLXKDVL 352                         |
| Qy   | 482 ALVGAVTSLQQELAHMRA-----RTHAPYGPYGPYHPPHAD 520                      | Qy | 482 ALVGAVTSLQQELAHMRA-----RTHAPYGPYGPYHPPHAD 520                      |
| Db   | 353 NLTKNMSQJQDDLKDQKOAASSQSPTRIIPQQFSSSSYFQDGHVPFRYGPYGIQKNDH-- 410   | Db | 353 NLTKNMSQJQDDLKDQKOAASSQSPTRIIPQQFSSSSYFQDGHVPFRYGPYGIQKNDH-- 410   |
| Qy   | 521 TETPAQPPRYPAKAVYLPPPHIAPPGPPLSGAVPPSYPPVAVTPG-PAPPLHQHOPP-- 576    | Qy | 521 TETPAQPPRYPAKAVYLPPPHIAPPGPPLSGAVPPSYPPVAVTPG-PAPPLHQHOPP-- 576    |
| Db   | 411 -----PPVcvSQQLPMQPLHvQOPPMQ 435                                    | Db | 411 -----PPVcvSQQLPMQPLHvQOPPMQ 435                                    |
| Qy   | 432 HEVEQPEYDGRDDEPDRDFPYGEARPEPRPVDSRRAARQASGPH-----E 478             | Qy | 432 HEVEQPEYDGRDDEPDRDFPYGEARPEPRPVDSRRAARQASGPH-----E 478             |



|       |  |     |                       |   |
|-------|--|-----|-----------------------|---|
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Db                    | 801 PPPPPP 805  |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 297 |                       |   |
| Qy    | 373 PHYPPPPAHPYPPGMLFAGPSPLEAQIAALVGAIADRQAGGLPAAGDHGTRGSAKRRH   | 432 |                       |   |
| Db    | 298 ELYPP--QPSGDNITGYMLPSGSYVAM-----FPSRN-----PACFQEQSPQDQAYMPY  | 334 |                       |   |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 |                       |   |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 |                       |   |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
| A     | Intonons: 623/3  |     |                       |   |
| A     | Note: F28A21.80  |     |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Query                 | 21 IYVAGF--LALYDSDGPGEALDPTDVTAAALPENPLPINVDHARCEVERVLAVNDP |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 334 | Match                 | 8.2%; Score 278; DB 2;                                      |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 | Best Local Similarity | 23.8%; Pred. No. 1.5e-07;                                   |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 | Matches               | 88; Mismatches 246; Indels 160; Gaps 32;                    |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
| A     | Intonons: 623/3  |     |                       |   |
| A     | Note: F28A21.80  |     |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Query Match           | 8.3%; Score 280; DB 2;                                      |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 334 | Best Local Similarity | 27.4%; Pred. No. 1.9e-07;                                   |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 | Matches               | 100; Mismatches 34; Indels 143; Gaps 16;                    |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 |                       |   |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
| A     | Intonons: 623/3  |     |                       |   |
| A     | Note: F28A21.80  |     |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Query Match           | 8.3%; Score 280; DB 2;                                      |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 334 | Best Local Similarity | 27.4%; Pred. No. 1.9e-07;                                   |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 | Matches               | 100; Mismatches 34; Indels 143; Gaps 16;                    |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 |                       |   |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
| A     | Intonons: 623/3  |     |                       |   |
| A     | Note: F28A21.80  |     |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Query Match           | 8.3%; Score 280; DB 2;                                      |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 334 | Best Local Similarity | 27.4%; Pred. No. 1.9e-07;                                   |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 | Matches               | 100; Mismatches 34; Indels 143; Gaps 16;                    |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 |                       |   |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
| A     | Intonons: 623/3  |     |                       |   |
| A     | Note: F28A21.80  |     |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Query Match           | 8.3%; Score 280; DB 2;                                      |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 334 | Best Local Similarity | 27.4%; Pred. No. 1.9e-07;                                   |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 | Matches               | 100; Mismatches 34; Indels 143; Gaps 16;                    |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 |                       |   |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
| A     | Intonons: 623/3  |     |                       |   |
| A     | Note: F28A21.80  |     |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Query Match           | 8.3%; Score 280; DB 2;                                      |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 334 | Best Local Similarity | 27.4%; Pred. No. 1.9e-07;                                   |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 | Matches               | 100; Mismatches 34; Indels 143; Gaps 16;                    |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 |                       |   |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
| A     | Intonons: 623/3  |     |                       |   |
| A     | Note: F28A21.80  |     |                       |   |
| <hr/> |  |     |                       |   |
| Qy    | 319 APPPGDGSSYLMIPASHY----NOLVTGQSAPRHPPLTACGLPAAGTVAVGHPGA-GPS  | 372 | Query Match           | 8.3%; Score 280; DB 2;                                      |
| Db    | 249 -VPQGSDDLISI:V:Q:K:PK:PSLMSLQNNLDNFKQHPR-----PACFQEQSPQDQAYMPY   | 334 | Best Local Similarity | 27.4%; Pred. No. 1.9e-07;                                   |
| Qy    | 433 EYBQPEYDCGDRDPRDPYPPGCEARPRPRPVDSSRAAQSAGPHETITALGAVTSIQLQ   | 492 | Matches               | 100; Mismatches 34; Indels 143; Gaps 16;                    |
| Db    | 335 DFD---DC-----VPPGE-----SSLYKDVLNLTKNISQLOQ   | 363 |                       |   |
| Qy    | 493 ELAHMRARTHAHYGPPVPPGVPHH--PHADTETPAQPRTPAKAVLPPPHIAPPGPPPL   | 549 |                       |   |
| Db    | 364 DKLKLQPAHINQPNRPP---HHSNPLDPGHSYFYRPPYQGKPKDHLQPLACY   | 419 |                       |   |
| Qy    | 550 SGAVPPPSYPPVATPGPAPPLHQPSAHAHPPPPGPTPPAASLPPQDPEAPGAEAAGAL   | 609 |                       |   |
| Db    | 420 QQA-----PVVQPNYA--PPTEGASNEAPKPKSVQEP-----449  |     |                       |   |
| Qy    | 610 VNASSAAHVNVDTARAAD   | 627 |                       |   |
| Db    | 450 -----VHIDASFAQD  | 459 |                       |   |
| Qy    | TO4859 extensin homolog F28A21.80 - Arabidopsis thaliana (mouse-ear cress)   | 4   |                       |   |
| C     | Species: Arabidopsis thaliana (mouse-ear cress)  |     |                       |   |
| C     | Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004  |     |                       |   |
| C     | Accession: To1859  |     |                       |   |
| R     | Bevan, M.; Mueller, M.W.; Muellein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer, R; Submitted to the Protein Sequence Database, February 1999 |     |                       |   |
| A     | Reference number: 215387   |     |                       |   |
| A     | Accession: To1859  |     |                       |   |
| A     | Molecule type: DNA   |     |                       |   |
| A     | Residues: 1-839 <BEV>  |     |                       |   |
| A     | Cross-references: UNIPROT:Q9SN46; EMBL:AL035526  |     |                       |   |
| A     | Experimental source: cultivar Columbia, BAC clone F28A21   |     |                       |   |
| C     | Genetics:  |     |                       |   |
| A     | Map position: 4  |     |                       |   |
|       |  |     |                       |   |

|  |  |
|--|--|
| Scoring table:                                     | BLOSUM62   |
| Gapop:   | 10.0 , Gapext: 0.5   |
| Searched:  | 1825181 seqs, 575374646 residues   |
| Total number of hits satisfying chosen parameters: | 1825181.   |
| Minimum DB seq length:                             | 0  |
| Maximum DB seq length:                             | 2000000000   |
| Post-processing:                                   | Minimum Match 0\$ Maximum Match 100\$  |
| Database :   | UniProt 02-*<br>1: uniprot_sprot:*   |
|  | 2: uniprot_trembl:*  |
| Pred. No.  | is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution. |
| SUMMARIES  |  |
| Result No.   | Query Score Match Length DB ID Description   |
| 1  | 3370 5 100.0 637 2 P89449 human herpesvirus 2  |
| 2  | 3370 5 99.6 638 2 Q69527 human herpesvirus 2   |
| 3  | 3351.5 99.0 636 2 Q69104 human herpesvirus 2   |
| 4  | 2334 69.0 635 2 Q69087 human herpesvirus 2   |
| 5  | 2334 68.9 635 1 VP40_HHV11   |
| 6  | 1820 53.8 329 1 P90341   |
| 7  | 1729 55.1 608 2 Q806A3   |
| 8  | 1368 50.4 651 2 Q8UY97   |
| 9  | 1361 40.2 562 2 Q918P4   |
| 10   | 1131 33.4 247 2 O12271 bovine herpesvirus 1  |
| 11   | 1079 5 31.9 646 1 VP40_EHV1B   |
| 12   | 1079 5 31.9 646 1 Q6S6T7   |
| 13   | 1079 5 31.9 646 2 AAS45919   |
| 14   | 1056 31.2 619 2 Q6X238   |
| 15   | 1056 31.2 619 2 AAR86135   |
| 16   | 1056 31.2 647 2 O39278   |
| 17   | 1010 29.8 621 1 VP40_BHV1C   |
| 18   | 1010 29.8 621 2 Q76PK0   |
| 19   | 1010 29.8 621 2 CAB01600   |
| 20   | 960 28.4 524 2 Q83417  |
| 21   | 904 26.7 526 2 Q85027  |
| 22   | 838 5 24.8 639 2 Q78283  |
| 23   | 838 5 24.8 639 2 Q785G2  |
| 24   | 838 5 24.8 639 2 Q9Y2Z6  |
| 25   | 824 5 24.4 605 1 VP40_VZVD   |
| 26   | 824 5 24.4 605 2 Q6QCM2  |
| 27   | 824 5 24.4 605 2 AAT0715   |
| 28   | 824 5 24.4 605 2 AAT07791  |
| 29   | 792 23.4 643 2 Q9DPRI  |
| 30   | 787 23.2 663 2 Q9E6Z2  |
| 31   | 787 23.2 663 2 AAS01665  |

## ALIGNMENTS

|       |   |   |
|-------|---|---|
| Total | 1 | RESULT 1  |
|       |   | P89449 ID PRELIMINARY;  |
|       |   | AC P89449; PRT; 637 AA.   |
|       |   | DT 01-MAY-1997 (TREMBLrel. 03, Created)   |
|       |   | DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  |
|       |   | DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  |
|       |   | DE Release;   |
|       |   | GN Name=UL26;   |
|       |   | OS Human herpesvirus 2 (HHV-2) (Human herpes simplex virus 2).  |
|       |   | OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;   |
|       |   | OC Alphaherpesvirinae; Simplexvirus.  |
|       |   | NCBI_TaxID=10310; OX [1].   |
|       |   | RN RP SEQUENCE FROM N.A.  |
|       |   | RC STRAIN=HG52;   |
|       |   | RX MEDLINE=87111457; PubMed=3027242;  |
|       |   | RA McGeoch D.J., Moss H.W., McNab D., Frame M.C.;   |
|       |   | RT "DNA sequence and genetic content of the HindIII 1 region in the short unique component of the herpes simplex virus type 2 genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons."; RT J. Gen. Virol. 68:19-38(1987). |
|       |   | RT RN [2].  |
|       |   | RP SEQUENCE FROM N.A.   |
|       |   | RC STRAIN=HG52;   |
|       |   | RX MEDLINE=90228430; PubMed=2161906;  |
|       |   | RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A., Everett R., Fenwick M.;  |
|       |   | RT "Comparative DNA sequence analysis of the host shutoff genes of different strains of herpes simplex virus: type 2 strain HG52 encodes a truncated UL41 product"; RT J. Gen. Virol. 71:1387-1390(1990).   |
|       |   | RL RN [3].  |
|       |   | RP SEQUENCE FROM N.A.   |
|       |   | RC STRAIN=HG52;   |
|       |   | RX MEDLINE=92113549; PubMed=1662697;  |
|       |   | RA McGeoch D.J., Cunningham C., McIntyre G., Dolan A., Everett R., Fenwick M.;  |
|       |   | RT "Comparative sequence analysis of the long unique regions in the genomes of herpes simplex viruses types 1 and 2"; RT J. Gen. Virol. 72:3057-3075(1991).   |
|       |   | RL RN [4].  |
|       |   | RP SEQUENCE FROM N.A.   |
|       |   | RC STRAIN=HG52;   |
|       |   | RX MEDLINE=92356101; PubMed=1322965;  |
|       |   | RA Barnett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;  |
|       |   | RT "A novel herpes simplex virus gene (UL49A) encodes a putative membrane protein with counterparts in other herpesviruses."; RT J. Gen. Virol. 73:2167-2171(1992).   |
|       |   | RL RN [5].  |
|       |   | RP SEQUENCE FROM N.A.   |
|       |   | RC STRAIN=HG52;   |
|       |   | RA Dolan A.;  |
|       |   | RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  |

|            |   |                       |  |
|------------|---|-----------------------|--|
| DR         | EMBL; Z86099; CAB06750.1; -.  | RN                    | [1] SEQUENCE FROM N.A.   |
| DR         | HSSP; Q69527; 1AT3.   | RP                    | O'Shaanessy D.J., O'Donnell K.C., Hellmig B., Shabon U., O'Brien S., Debouch C.M., |
| DR         | MEROS; S21; 001; -.   | RA                    | Chantrelain P., Holmes S., Abdell-Meguid S.S.,                                     |
| DR         | GO; GO:0008233; F:peptidase activity; IEA.                          | RA                    | Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.                            |
| DR         | GO; GO:0004252; F:serine-type endopeptidase activity; IEA.          | RL                    | EMBL; U4329; AAA92139; 1; -.   |
| DR         | GO; GO:0005508; P:proteolysis and peptidolysis; IEA.                | DR                    | PDB; 1AT3; X-ray; A/B=1-247.   |
| DR         | InterPro; IPR001847; Peptidase S21.                                 | DR                    | MEROPS; S21-001; -.  |
| DR         | PF00716; Peptidase S21; 1.  | DR                    | GO; GO:0008233; F:peptidase activity; IEA.   |
| DR         | PRINTS; PR00236; HSVCAPSTD40.                                       | DR                    | GO; GO:0004252; F:serine-type endopeptidase activity; IEA.                         |
| RW         | Protease.   | GO                    | GO:0006508; P:proteolysis and peptidolysis; IEA.                                   |
| SEQUENCE   | 637 AA; 66940 MW; 96B9D1F68C8DAB9C CRC64;                           | DR                    | InterPro; IPR001847; Peptidase S21.  |
| Qy         | Query Match Score 3385; DB 2; Length 637;                           | DR                    | PRINTS; PR00236; HSVCAPSTD40.  |
| Db         | Best Local Similarity 100.0%; Pred. No. 4-2e-145;                   | KW                    | Protease.  |
| Matches    | 637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;                | SO                    | SEQUENCE 638 AA; 67028 MW; 992F230937BBAE89 CRC64;                                 |
| Qy         | 1 MASAEMMERLEAPLDPRAPIVAGFLALYDSDGPGLADPDTYRALPENPLPINVD 60         | Query Match           | 99.6%; Score 3370.5; DB 2; Length 638;   |
| Db         | 1 MASAEMMERLEAPLDPRAPIVAGFLALYDSDGPGLADPDTYRALPENPLPINVD 60         | Best Local Similarity | 99.7%; Pred. No. 1.9e-144;   |
| Matches    | 636; Conservative 1; Mismatches 0; Indels 0; Gaps 1;                | Matches               | 636; Conservative 1; Mismatches 0; Indels 1; Gaps 1;                               |
| Qy         | 61 HRARCEVGRVLAVNDPRGFPLDPTYRALPENPLPINVD 60                        | Qy                    | 1 MASAEMMERLEAPLDPRAPIVAGFLALYDSDGPGLADPDTYRALPENPLPINVD 60                        |
| Db         | 61 HRARCEVGRVLAVNDPRGFPLDPTYRALPENPLPINVD 60                        | Db                    | 1 MASAEMMERLEAPLDPRAPIVAGFLALYDSDGPGLADPDTYRALPENPLPINVD 60                        |
| Qy         | 121 ITNLYPSVLSPIRKRGDDEVPPRTLFAHALCAIGRLGTIVTDTSLDAIAJAPFHLD 180    | Qy                    | 61 HRARCEVGRVLAVNDPRGFPLDPTYRALPENPLPINVD 60                                       |
| Db         | 121 ITNLYPSVLSPIRKRGDDEVPPRTLFAHALCAIGRLGTIVTDTSLDAIAJAPFHLD 180    | Db                    | 61 HRARCEVGRVLAVNDPRGFPLDPTYRALPENPLPINVD 60                                       |
| Qy         | 181 ATREGVREAAEELALAGRTWAPGVEALTHILLSTAVNNNMRLDWSLVAERRQAGIA 240    | Qy                    | 121 ITNLYPSVLSISSLTKRGDDEVPPRTLFAHALCAIGRLGTIVTDTSLDAIAJAPFHLD 180                 |
| Db         | 181 ATREGVREAAEELALAGRTWAPGVEALTHILLSTAVNNNMRLDWSLVAERRQAGIA 240    | Db                    | 121 ITNLYPSVLSISSLTKRGDDEVPPRTLFAHALCAIGRLGTIVTDTSLDAIAJAPFHLD 180                 |
| Qy         | 241 GHTYLOASEKPKIWGAEASAPABERGKTKGAFGAMDTSPASVAPQAVADQAVASSSS 300   | Qy                    | 181 ATREGVREAAEELAAGRTWAPGVEALTHILLSTAVNNNMRLDWSLVAERRQAGIA 240                    |
| Db         | 241 GHTYLOASEKPKIWGAEASAPABERGKTKGAFGAMDTSPASVAPQAVADQAVASSSS 300   | Db                    | 181 ATREGVREAAEELAAGRTWAPGVEALTHILLSTAVNNNMRLDWSLVAERRQAGIA 240                    |
| Qy         | 301 SSSPPAPADMNPV/SAGGAPPPPGDSYLWTPASHNQLVTGOSAPRHPITAGLPAAG 360    | Qy                    | 241 GHTYLOASEKPKIWGAEASAPABERGKTKGAFGAMDTSPASVAPQAVADQAVASSSS 300                  |
| Db         | 301 SSSPPAPADMNPV/SAGGAPPPPGDSYLWTPASHNQLVTGOSAPRHPITAGLPAAG 360    | Db                    | 241 GHTYLOASEKPKIWGAEASAPABERGKTKGAFGAMDTSPASVAPQAVADQAVASSSS 300                  |
| Qy         | 361 TVAYGHPCAGSPSPHYPPPAHPYPCMLFAGPSPLEAQIAALVGIAADRQAGGLPAAAG 420  | Qy                    | 360 GTVAYGHPCAGSPSPHYPPPAHPYPCMLFAGPSPLEAQIAALVGIAADRQAGGLPAAAG 419                |
| Db         | 361 TVAYGHPCAGSPSPHYPPPAHPYPCMLFAGPSPLEAQIAALVGIAADRQAGGLPAAAG 420  | Db                    | 361 GTVAYGHPCAGSPSPHYPPPAHPYPCMLFAGPSPLEAQIAALVGIAADRQAGGLPAAAG 420                |
| Qy         | 421 HGTGSAKRYRREVEQDPYDCGDBDEPDRDPFYPGEARPEPDRPVDSSRAAQASGPHETI 480 | Qy                    | 300 SSSFPAPADMNPV/SAGGAPPPPGDSYLWTPASHNQLVTGOSAPRHPITAGLPAAG 359                   |
| Db         | 421 HGTGSAKRYRREVEQDPYDCGDBDEPDRDPFYPGEARPEPDRPVDSSRAAQASGPHETI 480 | Db                    | 301 SSSFPAPADMNPV/SAGGAPPPPGDSYLWTPASHNQLVTGOSAPRHPITAGLPAAG 360                   |
| Qy         | 481 TALVGAVTSQQLAHMRAHTAIGYGPYPVCPYHPHADETPAOPPRYPAKAVLPP 540       | Qy                    | 420 DHGIRGSAKRREVEQEPYDGRDPPDREPPYYPEARPPRPPVDSRRAQRQASCPHET 479                   |
| Db         | 481 TALVGAVTSQQLAHMRAHTAIGYGPYPVCPYHPHADETPAOPPRYPAKAVLPP 540       | Db                    | 421 DHGIRGSAKRREVEQEPYDGRDPPDREPPYYPEARPPRPPVDSRRAQRQASCPHET 480                   |
| Qy         | 541 HIAPPGPPLSGAVPPPSYPPVAVTPGPAPLHQPSPAHAPPPIPQPTPPAASLQPQE 600    | Qy                    | 480 ITALVGAVTSQQLAHMRAHTAIGYGPYPVCPYHPHADETPAOPPRYPAKAVLPP 539                     |
| Db         | 541 HIAPPGPPLSGAVPPPSYPPVAVTPGPAPLHQPSPAHAPPPIPQPTPPAASLQPQE 600    | Db                    | 481 ITALVGAVTSQQLAHMRAHTAIGYGPYPVCPYHPHADETPAOPPRYPAKAVLPP 540                     |
| Qy         | 601 ACPGAAGALVNSAANNDTARAADLFVSMQMGSR 637                           | Qy                    | 540 PHIAPCPPLSGAVPPPSYPPVAVTPGPAPLHQPSPAHAPPPIPQPTPPAASLQPQE 599                   |
| Db         | 601 ACPGAAGALVNSAANNDTARAADLFVSMQMGSR 637                           | Db                    | 541 PHIAPCPPLSGAVPPPSYPPVAVTPGPAPLHQPSPAHAPPPIPQPTPPAASLQPQE 600                   |
| Qy         | 600 BAPGAAGALVNSAANNDTARAADLFVSMQMGSR 637                           | Qy                    | 600 BAPGAAGALVNSAANNDTARAADLFVSMQMGSR 637  |
| Q69527     | PRELIMINARY;  | PRT;                  | 638 AA.  |
| AC         | 069527;   | RESULT 3              | 601 EAPGAAGALVNSAANNDTARAADLFVSMQMGSR 638  |
| DT         | 01-NOV-1996 (TREMBLrel. 01, Created)                                | Q69104                | PRELIMINARY;   |
| DT         | 01-NOV-1996 (TREMBLrel. 01, Last sequence update)                   | ID                    | PRT; 636 AA.   |
| DT         | 01-OCT-2003 (TREMBLrel. 25, Last annotation update)                 | AC                    | 069104;  |
| DB         | U126 Protease.  | DT                    | 01-NOV-1996 (TREMBLrel. 01, Created)   |
| OS         | Human herpesvirus 2 (HHV-2) (Human herpes simplex virus 2).         | DT                    | 01-NOV-1996 (TREMBLrel. 01, Last sequence update)                                  |
| OC         | Viruses; dsDNA viruses, no RNA stage; Herpesviridae;                | DT                    | 01-OCT-2003 (TREMBLrel. 25, Last annotation update)                                |
| NCBI_TaxID | =10310;   | DT                    |  |

| Protease (Fragment).   |  | Name=ICP35;   |  |
|--|--|---|--|
| GN Human herpesvirus 2 (HHV-2) (Human herpes simplex virus 2).               |  | OS  |  |
| OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;                      |  | OC  |  |
| OC Alphaherpesvirinae; Simplexvirus.   |  | OC  |  |
| NCBI_TaxID=10310;  |  | RN [1]  |  |
| SPTRAIN=G.   |  | SPSEQUENCE FROM N.A.  |  |
| RA Steffy K.K.S.; Schoen S.; Chen C.-M.;                                     |  | RT "Nucleotide sequence of the herpes simplex virus type 2 gene encoding the protease and capsid protein ICP35."; |  |
| RL J. Gen. Virol. 0:0-0(1995).   |  | RL GO:0008233; F:peptidase activity; IEA.   |  |
| DR GO:0006508; P:proteinase; P:peptidase activity; IEA.                      |  | DR GO:0008232; F:serine-type endopeptidase activity; IEA.   |  |
| DR InterPro; IPR01847; Peptidase_S21.  |  | DR InterPro; IPR01847; Peptidase_S21.   |  |
| DR PRINTS; PR00236; Peptidase_S21; 1.  |  | DR PRINTS; PR00236; HSVCAPSIDP40.   |  |
| KW Protease.   |  | KW NON TER 1 1  |  |
| FT SQ 636 AA; 66785 MW; 817FF5814F7A48C3 CRC64;                              |  | FT SQ 636 AA; 66785 MW; 817FF5814F7A48C3 CRC64;   |  |
| Query Match 99.0%; Score 3351.5; DB 2; Length 636;                           |  | Query Match 99.0%; Score 3351.5; DB 2; Length 636;  |  |
| Best Local Similarity 99.4%; Pred. No. 1..e-143;                             |  | Best Local Similarity 99.4%; Pred. No. 1..e-143;  |  |
| Matches 633; Conservative 1; Mismatches 2; Indels 1; Gaps 1;                 |  | Matches 633; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  |  |
| Qy 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60 |  | Qy 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60                                      |  |
| Db 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60 |  | Db 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60                                      |  |
| Qy 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60      |  | Qy 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60   |  |
| Db 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60      |  | Db 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60   |  |
| Qy 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180     |  | Qy 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180  |  |
| Db 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180     |  | Db 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180  |  |
| Qy 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240       |  | Qy 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240  |  |
| Db 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240       |  | Db 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240  |  |
| Qy 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300       |  | Qy 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300  |  |
| Db 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300       |  | Db 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300  |  |
| Qy 361 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 420         |  | Qy 361 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 420  |  |
| Db 360 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 419         |  | Db 360 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 419  |  |
| Qy 421 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 480        |  | Qy 421 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 480   |  |
| Db 420 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 479        |  | Db 420 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 479   |  |
| Qy 481 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 600          |  | Qy 481 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 600   |  |
| Db 480 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 599          |  | Db 480 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 599   |  |
| Qy 541 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 600            |  | Qy 541 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 600   |  |
| Db 540 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 637            |  | Db 540 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 637   |  |
| Qy 601 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637                              |  | Qy 601 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637   |  |
| Db 600 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637                              |  | Db 600 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637   |  |
| RESULT 4   |  | RESULT 4  |  |
| Q69087 ID 069087 PRELIMINARY;  |  | Q69087 ID 069087 PRELIMINARY;   |  |
| AC DT 01-NOV-1996 (TREMBLrel. 01, Created)                                   |  | AC DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)   |  |
| DT DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)                    |  | DT DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)   |  |
| DE UL26.   |  | DE UL26.  |  |
| OS Human herpesvirus 1 (HHV-1) (Human herpes simplex virus 1).               |  | OS Human herpesvirus 1 (HHV-1) (Human herpes simplex virus 1).  |  |
| OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;                      |  | OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;   |  |
| OC Alphaherpesvirinae; Simplexvirus.   |  | OC Alphaherpesvirinae; Simplexvirus.  |  |
| NCBI_TaxID=10310;  |  | RN [1]  |  |
| SPTRAIN=G.   |  | SPSEQUENCE FROM N.A.  |  |
| RA Steffy K.K.S.; Schoen S.; Chen C.-M.;                                     |  | RT "Nucleotide sequence of the herpes simplex virus type 2 gene encoding the protease and capsid protein ICP35."; |  |
| RL J. Gen. Virol. 0:0-0(1995).   |  | RL GO:0008233; F:peptidase activity; IEA.   |  |
| DR GO:0006508; P:proteinase; P:peptidase activity; IEA.                      |  | DR GO:0008232; F:serine-type endopeptidase activity; IEA.   |  |
| DR InterPro; IPR01847; Peptidase_S21.  |  | DR InterPro; IPR01847; Peptidase_S21.   |  |
| DR PRINTS; PR00236; Peptidase_S21; 1.  |  | DR PRINTS; PR00236; HSVCAPSIDP40.   |  |
| KW Protease.   |  | KW NON TER 1 1  |  |
| FT SQ 636 AA; 66785 MW; 817FF5814F7A48C3 CRC64;                              |  | FT SQ 636 AA; 66785 MW; 817FF5814F7A48C3 CRC64;   |  |
| Query Match 99.0%; Score 3351.5; DB 2; Length 636;                           |  | Query Match 99.0%; Score 3351.5; DB 2; Length 636;  |  |
| Best Local Similarity 99.4%; Pred. No. 1..e-143;                             |  | Best Local Similarity 99.4%; Pred. No. 1..e-143;  |  |
| Matches 633; Conservative 1; Mismatches 2; Indels 1; Gaps 1;                 |  | Matches 633; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  |  |
| Qy 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60 |  | Qy 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60                                      |  |
| Db 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60 |  | Db 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60                                      |  |
| Qy 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60      |  | Qy 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60   |  |
| Db 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60      |  | Db 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60   |  |
| Qy 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180     |  | Qy 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180  |  |
| Db 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180     |  | Db 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180  |  |
| Qy 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240       |  | Qy 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240  |  |
| Db 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240       |  | Db 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240  |  |
| Qy 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300       |  | Qy 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300  |  |
| Db 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300       |  | Db 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300  |  |
| Qy 361 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 420         |  | Qy 361 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 420  |  |
| Db 360 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 419         |  | Db 360 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 419  |  |
| Qy 421 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 480        |  | Qy 421 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 480   |  |
| Db 420 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 479        |  | Db 420 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 479   |  |
| Qy 481 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 600          |  | Qy 481 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 600   |  |
| Db 480 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 599          |  | Db 480 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 599   |  |
| Qy 541 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 600            |  | Qy 541 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 600   |  |
| Db 540 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 637            |  | Db 540 HIAPPGPPLSGAVPPSVPPVAVTPGPAPLHPSPAHAPPPPBPPTPPAAASLQPB 637   |  |
| Qy 601 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637                              |  | Qy 601 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637   |  |
| Db 600 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637                              |  | Db 600 APGAEGALVNASSAHHNVDTARAIDLFLVSONGMGR 637   |  |
| RESULT 4   |  | RESULT 4  |  |
| Q69087 ID 069087 PRELIMINARY;  |  | Q69087 ID 069087 PRELIMINARY;   |  |
| AC DT 01-NOV-1996 (TREMBLrel. 01, Created)                                   |  | AC DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)   |  |
| DT DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)                    |  | DT DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)   |  |
| DE UL26.   |  | DE UL26.  |  |
| OS Human herpesvirus 1 (HHV-1) (Human herpes simplex virus 1).               |  | OS Human herpesvirus 1 (HHV-1) (Human herpes simplex virus 1).  |  |
| OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;                      |  | OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;   |  |
| OC Alphaherpesvirinae; Simplexvirus.   |  | OC Alphaherpesvirinae; Simplexvirus.  |  |
| NCBI_TaxID=10310;  |  | RN [1]  |  |
| SPTRAIN=G.   |  | SPSEQUENCE FROM N.A.  |  |
| RA Steffy K.K.S.; Schoen S.; Chen C.-M.;                                     |  | RT "Nucleotide sequence of the herpes simplex virus type 2 gene encoding the protease and capsid protein ICP35."; |  |
| RL J. Gen. Virol. 0:0-0(1995).   |  | RL GO:0008233; F:peptidase activity; IEA.   |  |
| DR GO:0006508; P:proteinase; P:peptidase activity; IEA.                      |  | DR GO:0008232; F:serine-type endopeptidase activity; IEA.   |  |
| DR InterPro; IPR01847; Peptidase_S21.  |  | DR InterPro; IPR01847; Peptidase_S21.   |  |
| DR PRINTS; PR00236; Peptidase_S21; 1.  |  | DR PRINTS; PR00236; HSVCAPSIDP40.   |  |
| KW Protease.   |  | KW NON TER 1 1  |  |
| FT SQ 636 AA; 66785 MW; 817FF5814F7A48C3 CRC64;                              |  | FT SQ 636 AA; 66785 MW; 817FF5814F7A48C3 CRC64;   |  |
| Query Match 99.0%; Score 3351.5; DB 2; Length 636;                           |  | Query Match 99.0%; Score 3351.5; DB 2; Length 636;  |  |
| Best Local Similarity 99.4%; Pred. No. 1..e-143;                             |  | Best Local Similarity 99.4%; Pred. No. 1..e-143;  |  |
| Matches 633; Conservative 1; Mismatches 2; Indels 1; Gaps 1;                 |  | Matches 633; Conservative 1; Mismatches 2; Indels 1; Gaps 1;  |  |
| Qy 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60 |  | Qy 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60                                      |  |
| Db 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60 |  | Db 1 MSAEMMRERLEAPLPDRAPVPIYAGFLALYQACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60                                      |  |
| Qy 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60      |  | Qy 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60   |  |
| Db 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60      |  | Db 61 HRARCEVGRVLAVVNDRPGPFVGLIACVLERVLETASAIAFERRGPALEPRLPENPLPINVD 60   |  |
| Qy 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180     |  | Qy 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180  |  |
| Db 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180     |  | Db 121 ITNYLPVSLSLKTRRGDDEVPPDRTLFAHVALCAIGRRGLGTIVYDTSLDAAIAPFRHLDP 180  |  |
| Qy 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240       |  | Qy 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240  |  |
| Db 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240       |  | Db 181 ATREGVRRREAAEELALLAGRITWAPGEALTHTLISTAVNNMLDRDLSLVAERRQAGIA 240  |  |
| Qy 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300       |  | Qy 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300  |  |
| Db 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300       |  | Db 241 GHTYQASEKFKIWIWAEASAPAPERGKTYGKTPGAGMDTSPASVPAQVAVRAQVASSSS 300  |  |
| Qy 361 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 420         |  | Qy 361 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 420  |  |
| Db 360 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 419         |  | Db 360 TVALYHGPAGKSPHYPPPAHPPGMLFAGPSPLEAQTAALVGAIAADRQAGGLPAAGD 419  |  |
| Qy 421 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 480        |  | Qy 421 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 480   |  |
| Db 420 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 479        |  | Db 420 HGIRGSAKRRRHEVEQPYDXGRDPEPDRDPYPPGEARPEPREVDSSRAARQASGHETI 479   |  |
| Qy 481 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 600          |  | Qy 481 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 600   |  |
| Db 480 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 599          |  | Db 480 TALVGAVTSLSQQLELAHLBARTHAPYGYPPVGPYHPPGTPPPGTPPPAAASLQPB 599   |  |
| Qy 541 HIAPP   |  |   |  |

|                    |   |   |             |   |
|--------------------|---|---|-------------|---|
| Db                 | 474   | TNETITALMGAIVTSLQQLELAMNMRARTSAPGMYNTPYAHYRPQYGEPEPTTHPALCPEA | 533         | FT PROTEP 7611 635 C-terminal peptide.                                |
| Db                 | 535   | YVLPPIPHIAPPGLPLSGA- YVPPPSYTPVAVTPGPAPLHQPSPAHPPPPGPTTPP     | 592         | FT DOMAIN 514 590 Pro-rich.   |
| Qy                 | 534   | YVRPPIPHSAPGPPQGASHATTPYAPACCPGPPP--DCPSTOTRPLPTEAPP          | 590         | FT SITE 247 248 Cleavage (by the protease).                           |
| Db                 | 593   | AASLQPEAPEAAGAAGLVNASSAAHVVDTAAADLFVSONGSR                    | 637         | FT SITE 610 611 Cleavage (by the protease).                           |
| Qy                 | 591   | AATGQSEASNAEAGALVNASSAAHVVDTAAADLFVSONGAR                     | 635         | FT ACT SITE 61 61 Charge relay system (By similarity).                |
| Db                 | 591   | AATGQSEASNAEAGALVNASSAAHVVDTAAADLFVSONGAR                     | 635         | FT ACT SITE 129 129 Charge relay system (By similarity).              |
| SO                 | SEQUENCE  | 635 AA;   | 66470 MW;   | F3B3CTD42F3D062D CRC64;   |
| Query Match        | Best Local Similarity   | 68.9%;  | Score 2333; | DB 1; Length 635;   |
| Matches            | 477;  | Conservative  | 30;         | Mismatches 120; Indels 18; Gaps 7;                                    |
| Qy                 | 1   | MASAEMRERLEAPLDRAYPIVYAGFLALYDSDGDPGELADPDPTVRAALPPENPLPINVD  | 60          | Db 1 MAADAGDGMEEPLPDRAYPIVYAGFLALYDSDGELADPDPTVRAALPPENPLPINVD 60     |
| Qy                 | 61  | HRADCEEVGRVLAVINDPROPFVGLIACQLERLVETAAASANIFERGRPLSREBERTYL   | 120         | Db 61 HRADCEEVGRVLAVINDPROPFVGLIACQLERLVETAAASANIFERGRPLSREBERTYL 120 |
| Qy                 | 61  | HRASCEEVGRVLAVVDDPRGFFVGLIACQLERLVETAAASAAIFERGRPLSREBERTYL   | 120         | Db 61 HRASCEEVGRVLAVVDDPRGFFVGLIACQLERLVETAAASAAIFERGRPLSREBERTYL 120 |
| Qy                 | 121   | ITNVLPSVSLSTKRGDEVPDRTLFAHYALCAIGRLGTIVTYDLSDAIAJAPFHLDP      | 180         | Db 121 ITNVLPSVSLSTKRGDEVPDRTLFAHYALCAIGRLGTIVTYDLSDAIAJAPFHLDP 180   |
| Qy                 | 121   | ITNVLPSVSLATRLLGEAHFDRLTIAHYALCAIGRLGTIVTYDLSDAIAJAPFHLDSP    | 180         | Db 121 ITNVLPSVSLATRLLGEAHFDRLTIAHYALCAIGRLGTIVTYDLSDAIAJAPFHLDSP 180 |
| Qy                 | 181   | ATREGVRRRAAAEELAALAGTWAQGEVALTHLILSTAVNNMLDRWSLVAERRRAGIA     | 240         | Db 181 ASRERARLAAEELAALSRTWAQEVALTHLILSTAVNNMLDRWSLVAERRRAGIA 240     |
| Qy                 | 241   | GHTYLIQASEKFKingEASAPERCYKTAGPMGDTSPASVPAQD--VAVRAQVAS        | 296         | Db 241 GHTYLIQASEKFKingEASAPERCYKTAGPMGDTSPASVPAQD--VAVRAQVAS 296     |
| Qy                 | 241   | GHTYLIQASEKFKingEASAPERCYKTAGPMGDTSPASVPAQD--VAVRAQVAS        | 296         | Db 241 GHTYLIQASEKFKingEASAPERCYKTAGPMGDTSPASVPAQD--VAVRAQVAS 296     |
| Qy                 | 297   | SSSSSFEPADAMNFSVASSAGAAPPGDSYSLWTFASHYQNOLVGTQSAPPHPPLTAGL    | 356         | Db 297 SSSSSFEPADAMNFSVASSAGAAPPGDSYSLWTFASHYQNOLVGTQSAPPHPPLTAGL 356 |
| Qy                 | 301   | S-----PVLPPMNPVPTSTPAPPGDSYSLWTFASHYQNOLVGTQSAPPHPPLTAGF      | 354         | Db 301 S-----PVLPPMNPVPTSTPAPPGDSYSLWTFASHYQNOLVGTQSAPPHPPLTAGF 354   |
| Qy                 | 357   | P-AAGTVAYGHPGAGGSPSPHYPYPPHYPYPPHYPYPPHYPYPPHYPYPPHYPYPPH     | 415         | Db 357 P-AAGTVAYGHPGAGGSPSPHYPYPPHYPYPPHYPYPPHYPYPPHYPYPPHYPYPPH      |
| Qy                 | 355   | PAAAQSVAYGHPGAGSQQHYPHYAHQPVGLFSQSPLEQIAAVGTAADRGQGP          | 414         | Db 355 PAAAQSVAYGHPGAGSQQHYPHYAHQPVGLFSQSPLEQIAAVGTAADRGQGP           |
| Qy                 | 416   | AAAGGHHGIGRSARRRHEVQPEYDQGRDFDPPYGPGEARPERPPDSRRAQASG         | 475         | Db 416 AAAGGHHGIGRSARRRHEVQPEYDQGRDFDPPYGPGEARPERPPDSRRAQASG 475      |
| Qy                 | 415   | -AAGDPGVRSQGSKRREYAGPSESTCDQDDEPDADYFYPGEARGARGVDRRAAHSQPG    | 473         | Db 415 -AAGDPGVRSQGSKRREYAGPSESTCDQDDEPDADYFYPGEARGARGVDRRAAHSQPG 473 |
| Qy                 | 476   | PHETITALVGAVTSIQQELAHMRARTHAPYGPYPYHPPHADTE-TPAQPPRPPAKA      | 534         | Db 476 PHETITALVGAVTSIQQELAHMRARTHAPYGPYPYHPPHADTE-TPAQPPRPPAKA 534   |
| Qy                 | 474   | TNETITALMVAIQLQSQELAHMRARTHAPYGPYHPPHADTE-TPAQPPRPPAKA        | 533         | Db 474 TNETITALMVAIQLQSQELAHMRARTHAPYGPYHPPHADTE-TPAQPPRPPAKA 533     |
| Qy                 | 535   | VYLPPPHIAPPGLPLSGA- VPPPSYPPVAVTPGPAPLHQPSPAHAPPPTPPTPPP      | 592         | Db 535 VYLPPPHIAPPGLPLSGA- VPPPSYPPVAVTPGPAPLHQPSPAHAPPPTPPTPPP 592   |
| Qy                 | 534   | YVRPHPSAYGPQCPASHAATPPVAPAACTPPGPPB--PCPSQTQTRALPTEAPP        | 590         | Db 534 YVRPHPSAYGPQCPASHAATPPVAPAACTPPGPPB--PCPSQTQTRALPTEAPP 590     |
| Db                 | 593   | AASLQPEAPEAAGAAGLVNASSAAHVVDTAAADLFVSONGSR                    | 637         | Db 593 AASLQPEAPEAAGAAGLVNASSAAHVVDTAAADLFVSONGSR 637                 |
| Db                 | 591   | AATGQSEASNAEAGALVNASSAAHVVDTAAADLFVSONGAR                     | 635         | Db 591 AATGQSEASNAEAGALVNASSAAHVVDTAAADLFVSONGAR 635                  |
| RESULT             | 6   |   |             |   |
| ID                 | P90341  | PRELIMINARY;  | PRT;        | 329 AA.   |
| AC                 | P90341;   |   |             |   |
| DT                 | 01-MAY-1997   | (TREMBLrel. 03, Created)                                      |             |   |
| DT                 | 01-MAY-1997   | (TREMBLrel. 03, Last sequence update)                         |             |   |
| DT                 | 01-MAR-2004   | (TREMBLrel. 26, Last annotation update)                       |             |   |
| DB                 | UL2-5   | 5 protein.  |             |   |
| GN                 | Name-UL26-5;  |   |             |   |
| OS                 | Human herpesvirus 2 (HHV-2) (Human herpes simplex virus 2).   |   |             |   |
| OC                 | Alphaherpesviruses; Simplexvirus;   |   |             |   |
| NCBI_TaxID         | 10310;  |   |             |   |
| NCBI_TaxID         | 11  |   |             |   |
| SEQUENCE FROM N.A. |   |   |             |   |
| RP                 | SEQUENCE FROM N.A.  |   |             |   |
| RX                 | MEDLINE=81274327; PubMed=2839594;   |   |             |   |
| RA                 | McGeoch D.J., Dairymple M.A., Davison A.J., Frame M.C.,   |   |             |   |
| RA                 | Davison M.D., Rixon F.J., Davison A.J., Frame M.C.,   |   |             |   |
| RA                 | McNab D., Perry L.J., Scott J.E., Taylor P.J.,  |   |             |   |
| RA                 | "The complete DNA sequence of the long unique region in the genome of   |   |             |   |
| RT                 | "The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1."   |   |             |   |
| RT                 | herpes simplex virus type 1."   |   |             |   |
| RL                 | J. Gen. Virol. 69:1531-1574 (1988).   |   |             |   |
| RN                 | [2]   |   |             |   |
| RP                 | SEQUENCE OF 11-29; 77-91 AND 223-241.   |   |             |   |
| RX                 | MEDLINE=901027; PubMed=1328483;   |   |             |   |
| RX                 | Davison M.D., Rixon F.J., Davison A.J., Frame M.C.,   |   |             |   |
| RA                 | Davison M.D., Rixon F.J., Davison A.J., Frame M.C.,   |   |             |   |
| RA                 | "Identification of genes encoding two capsid proteins (VP24 and VP26)   |   |             |   |
| RT                 | "Identification of genes encoding two capsid proteins (VP24 and VP26)   |   |             |   |
| RT                 | of herpes simplex virus type 1."  |   |             |   |
| RL                 | J. Gen. Virol. 73:2709-2713 (1992).   |   |             |   |
| CC                 | -I- FUNCTION: VP22A is a component of the capsid core involved in processing and packaging of progeny DNA. VP24 is a protease which can proteolytically cleave itself and VP22A at the C-terminus.  |   |             |   |
| CC                 | -I- CATALYTIC ACTIVITY: Cleaves -Ala-1-Ser- and -Ala-1-Ala- bonds in the capsid protein.  |   |             |   |
| CC                 | -I- PTM: VP22A is phosphorylated.   |   |             |   |
| CC                 | -I- SIMILARITY: Belongs to peptidase family S21.  |   |             |   |
| CC                 | -----   |   |             |   |
| CC                 | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a> or send an email to license@isb-sib.ch). |   |             |   |
| CC                 | -----   |   |             |   |
| DR                 | EMBL; X14112; CAA32318.1; -.  |   |             |   |
| DR                 | DR  | PRELIMINARY;  | PRT;        | 329 AA.   |
| DR                 | PIR; H30084; WMBW6.   |   |             |   |
| DR                 | HSSP; Q69527; 1AT3.   |   |             |   |
| DR                 | MEROPS; S21_001; -.   |   |             |   |
| DR                 | InterPro; IPR001847; Peptidase_S21.   |   |             |   |
| DR                 | PFAM; PF00716; Peptidase_S21; 1.  |   |             |   |
| DR                 | PRINTS; PR00236; HSCVCAPISTD40.   |   |             |   |
| KW                 | Capsid assembly; Coat protein; Direct protein sequencing; Hydrolase; phosphorylation; Serine protease.  |   |             |   |
| KW                 | CHAIN 1 635 Gene UI26 protein.  |   |             |   |
| PT                 | CHAIN 1 635 UL26.5 protein.   |   |             |   |
| PT                 | CHAIN 1 2610 Capsid protein VP22A.  |   |             |   |
| PT                 | CHAIN 1 2610 Capsid protein VP22A.  |   |             |   |
| PT                 | CHAIN 1 2610 Capsid protein VP22A.  |   |             |   |

|        |   |              |     |         |
|--------|---|--------------|-----|---------|
| RC     | STRAIN=HG52;  | PRELIMINARY; | PT; | 608 AA. |
| RX     | MEDLINE=87111457; PubMed=3027242;   |              |     |         |
| RA     | McGeoch D.J., Moss H.W., McLab D., Frame M.C.;  |              |     |         |
| RT     | "DNA sequence and genetic content of the HindIII 1 region in the short unique component of the herpes simplex virus type 2 genome: identification of the gene encoding glycoprotein G, and evolutionary comparisons"; |              |     |         |
| RT     | J. Gen. Viroi. 68:19-38(1987).  |              |     |         |
| RN     |   |              |     |         |
| RP     | SEQUENCE FROM N.A.  |              |     |         |
| RC     | STRAIN=HG52;  |              |     |         |
| RX     | MEDLINE=90278430; PubMed=2161906;   |              |     |         |
| RA     | Everett R., Fenwick M.;   |              |     |         |
| RT     | "Comparative DNA sequence analysis of the host shutoff genes of different strains of herpes simplex virus: type 2 strain HG52 encodes a truncated UL41 product.";   |              |     |         |
| RT     | J. Gen. Viroi. 71:1387-1390(1990).  |              |     |         |
| [3]    |   |              |     |         |
| RN     | SEQUENCE FROM N.A.  |              |     |         |
| RC     | STRAIN=HG52;  |              |     |         |
| RX     | MEDLINE=92113549; PubMed=1662697;   |              |     |         |
| RA     | McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;   |              |     |         |
| RT     | "Comparative sequence analysis of the long repeat regions and adjoining parts of the long unique regions in the genomes of herpes simplex viruses types 1 and 2.";  |              |     |         |
| RT     | J. Gen. Viroi. 72:3057-3075(1991).  |              |     |         |
| [4]    |   |              |     |         |
| RP     | SEQUENCE FROM N.A.  |              |     |         |
| RC     | STRAIN=HG52;  |              |     |         |
| RX     | MEDLINE=92356101; PubMed=1322965;   |              |     |         |
| RA     | Barnett B.C., Dolan A., Telford E.A.R., Davison A.J., McGeoch D.J.;   |              |     |         |
| RT     | "A novel herpes simplex virus gene (UL49A) encodes a putative membrane protein with counterparts in other herpesviruses.";  |              |     |         |
| RT     | J. Gen. Viroi. 73:2167-2171(1992).  |              |     |         |
| [5]    |   |              |     |         |
| RN     | SEQUENCE FROM N.A.  |              |     |         |
| RC     | STRAIN=HG52;  |              |     |         |
| RA     | Dolan A.;   |              |     |         |
| RL     | Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.   |              |     |         |
| DR     | EMBL: 266099; CAB6751.1; -;   |              |     |         |
| SQ     | SEQUENCE 329 AA; 33836 MW; 590B0428E63451E6 CRC64;  |              |     |         |
| Qy     | Query Match 53.8%; Score 1820; DB 2; Length 329; Best Local Similarity 100.0%; Pred. No. 8e-75; Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  |              |     |         |
| Db     | 1. MNPVASGAPAPPGDSYLWIPASHYHNLVTGQSAPRHPPLTACGLPAAGTVAYHGP 368  |              |     |         |
| Qy     | 309 MNPVASGAPAPPGDSYLWIPASHYHNLVTGQSAPRHPPLTACGLPAAGTVAYHGP 368   |              |     |         |
| Db     | 1. MNPVASGAPAPPGDSYLWIPASHYHNLVTGQSAPRHPPLTACGLPAAGTVAYHGP 60   |              |     |         |
| Qy     | 369 AGSPHYPAPPAHYPGMIFAGPSPLEACIAALVGATAADRGGLPAAGDHGIRSAA 428  |              |     |         |
| Db     | 61. AGSPHYPAPPAHYPGMIFAGPSPLEACIAALVGATAADRGGLPAAGDHGIRSAA 120  |              |     |         |
| Qy     | 309 RPRHEVEQPEYDCGRDEPDRDFPXTYPGERPPEPVDSRAARQASGPHEITIALVGA 488  |              |     |         |
| Db     | 121. RPRHEVEQPEYDCGRDEPDRDFPXTYPGERPPEPVDSRAARQASGPHEITIALVGA 180   |              |     |         |
| Qy     | 429 RPRHEVEQPEYDCGRDEPDRDFPXTYPGERPPEPVDSRAARQASGPHEITIALVGA 488  |              |     |         |
| Db     | 121. RPRHEVEQPEYDCGRDEPDRDFPXTYPGERPPEPVDSRAARQASGPHEITIALVGA 180   |              |     |         |
| Qy     | 489 SLQQELAHMRTHAPYKGPYPGVPGPYPHQPSPAHANPPPPGTPPPAASLQPPEAPGABA 548   |              |     |         |
| Db     | 181 SLQQELAHMRTHAPYKGPYPGVPGPYPHQPSPAHANPPPPGTPPPAASLQPPEAPGABA 240   |              |     |         |
| Qy     | 549 LSGAVPPPPYPPVAVTPGAPPLHQPSPAHNPQPPGTPPPAASLQPPEAPGABA 608   |              |     |         |
| Db     | 241 LSGAVPPPPYPPVAVTPGAPPLHQPSPAHNPQPPGTPPPAASLQPPEAPGABA 300   |              |     |         |
| Qy     | 609 LYNSAAAHNVDTARAADLFLVSMGMGSR 637  |              |     |         |
| Db     | 301 LYNSAAAHNVDTARAADLFLVSMGMGSR 329  |              |     |         |
| Qy     | 359 AGTVAYHGPAGSPHYPPPPAHGPSPGSPSPGSPSPGSPSPGSPSPGSPSPGSPSPGSPSP 345  |              |     |         |
| Db     | 346 P-PPAGAPAYPHHH----AGVYGVVFPGSPSPLEQIAALVGAADROQAGGLPAA 418  |              |     |         |
| Qy     | 419 GDHGIGRSAKRERRHEQPEYDCGRDEPDRDPPYYPGEARPEPRPV-DS---RRAARQA 473  |              |     |         |
| Q806A3 | RESULT 7  |              |     |         |

|                       |             |  |                  |   |
|-----------------------|-------------|--|------------------|---|
| Db                    | 397         | -----SGERRYEAER--DCDYDDRD-DAPYIFGEAARAPPRVYPPDSGGRGRRAAPSSA 445  | Qy               | 298 -----SSSSFPAPADNPVSASAGAPA-PPPGDGSYIWIWIPASHYNQL 339              |
| Qy                    | 474         | -SGPHETITALVGATSLQOELAHMARTAHPYG-PY-----PPVGPY----- 514  | Db               | 301 PSPPSPCDSPRIDGSSLVPSAOPMAYN-SPAGSPLPKPFQGMLFAGPSPLEAQ 359         |
| Db                    | 446         | : GTGASETIALVGATSLQOELAHMARA-APYQQYAAQPAQYAAQWE 504  | Qy               | 340 VTGO--SAPRHPPLTACGLPAAGTVAYGHPGAGSPSPHYPPPAHPYPPGMLFAGPSPLEAQ 397 |
| Qy                    | 515         | -----HHPHADTETPAQPRYPAKAVLVLPPHTAPPGPPLSGAVPPSYPPVAV-TP 565  | Db               | 360 VSSQDATAPLHPYIYAHQGVAG--YGPGVVPA--AGPSPLETQ 407                   |
| Db                    | 505         | APAGAALAHHP-----PPRH-----QPHATPANVPL--VPGQPCAYALAAQ 545  | Qy               | 398 IAAUQGIAADHQACCL---PAAGDHGIGRSKAKPRRHEVEQFED---CGDDEPDRD 449      |
| Qy                    | 566         | GPAPPHQPSPAHAPPPPPCPPTPPAASLDQPEAPGAEA-GALVNASSAAHVNVDTAR 624  | Db               | 408 IAALYGMATTAERQAGAALGSPASDSSAGLRSKHKRRCYERDYYGEFARGCSPP-- 464      |
| Db                    | 546         | : GPAGAAAPGGCA----PAAGAAAPTA----EAGETAVGALVNASSAATHVVDVGR 595  | Qy               | 450 FPYYPGB---ARPRPRPDSDR--RAAROASG-PH-ETTALVGATSLQOELAHMRAR 500      |
| Qy                    | 625         | ADLFVYQSQMMGSR 637   | Db               | 465 -PYYQGBYSGSGGGPGRRRRFFATRSAATPHSNETTAALVGAVASIQELTHLRSY 523       |
| Db                    | 596         | ADLFVYQSQMMGSR 608   | Qy               | 501 THAPYGPYPPVPGPYTHPHADTEPQQPRPAKAVYLPPPHIAPPCCPLSGAVPPPSYPP 560    |
| Qy                    | 561         | WAVTGPAPPPLHQSPAHAPHPHPPPPGPPPPPASL-----POPEAPGA----EAG 607  | Db               | 524 QHGAF-----VPSAAAQWTWPRPYFAP----AAAQASHQL 556                      |
| Qy                    | Q8JY97      | PRELIMINARY; PRT; 651 AA.  | Qy               | 557 QQAOPSCAPVTTQTPAQVPAALFAAPPAPVQSLGVYGAAPMEPRAGDAADVASEAD 616      |
| AC                    | Q8JY97;     |  | Db               |   |
| DT                    | 01-OCT-2002 | (TREMBLrel. 22, Created)   | Qy               | 608 A---LVNASSAAHVNVDTARAADLTVQSOMGSR 637                             |
| DT                    | 01-OCT-2002 | (TREMBLrel. 22, Last sequence update)  | Db               | 617 APPHINASCTTRVYDANRASDARAVAQMGDR 649                               |
| DT                    | 01-OCT-2003 | (TREMBLrel. 25, Last annotation update)  | DB               |   |
| DB                    |             | Protease.  | RESULT 9         |   |
| OS                    |             | Saimiriine herpesvirus 1 (strain MV-5-4-PSL) (Marmoset herpesviruses).   | Q918P4           |   |
| OC                    |             | Viruses; dsDNA viruses, no RNA stage; Herpesviridae;   | ID               |   |
| OC                    |             | Alphaherpesvirinae.  | AC               |   |
| OC                    |             | NCBI_TaxID:10355;  | DT               | Q918P4; PRELIMINARY; PRT; 562 AA.                                     |
| RN                    | [1]         |  | AC               | Q918P4; PRELIMINARY; PRT; 562 AA.                                     |
| RP                    |             | SEQUENCE FROM N.A.   | DT               | Q918P4; PRELIMINARY; PRT; 562 AA.                                     |
| RA                    |             | Bresnars M.A., Black D.H., Ritchey J.W., Eberle R.; Submitted (APR-07-2002) to the EMBL/GenBank/DBJ databases. | DT               | 01-DEC-2001 (TREMBLrel. 19, Created)                                  |
| RL                    |             | Submitted (APR-07-2002) to the EMBL/GenBank/DBJ databases.   | DT               | 01-DEC-2001 (TREMBLrel. 19, Last sequence update)                     |
| DR                    |             | EMBL; AY093366; AAM22798.1; -.   | DE               | 01-OCT-2003 (TREMBLrel. 25, Last annotation update)                   |
| DR                    |             | DR   | UL26             |   |
| DR                    |             | HSSP; Q69527; LAT3.  | DE               |   |
| DR                    |             | MEPROS; S21.001; -.  | DR               |   |
| DR                    |             | GO; GO:0008233; F: peptidase activity; IEA.  | DR               |   |
| DR                    |             | GO; GO:0002425; F: serine-type endopeptidase activity; IEA.  | DR               |   |
| DR                    |             | GO; GO:0006508; P: proteolysis and peptidolysis; IEA.  | DR               |   |
| DR                    |             | InterPro; IPR001847; Peptidase S21.  | DR               |   |
| DR                    |             | InterPro; IPR001847; Peptidase S21.  | DR               |   |
| DR                    |             | PRINTS; PRO0236; HSVCAPS1DP40.   | DR               |   |
| KW                    |             | Protease.  | SEQUENCE 562 AA; |   |
| SQ                    |             | 651 AA; 68104 MW; CP62E03E2EFP3BB9 CRC64;  | Q918P4           |   |
| Qy                    | 8           | ERLEAPLPDRAVP-IVYAGFLALYDSDGPGLALDPTWRAALPPNPINVDHARC 65   | Qy               | 59809 MW; 1E783D9743D90CBD CRC64;                                     |
| SEQUENCE              | 8           | ERLEAPLPDRAVP-IVYAGFLALYDSDGPGLALDPTWRAALPPNPINVDHARC 65   | Qy               | Query Match 40.4%; Score 1368.5; DB 2; Length 651;                    |
| Best Local Similarity | 49.1%       | Score 3.4e-54; Pred. No. 3.4e-54;  | Qy               | Best Local Similarity 51.1%; Pred. No. 6.5e-54;                       |
| Matches               | 340;        | Conservative 54; Mismatches 182; Indels 117; Gaps 25;  | Qy               | Matches 327; Conservative 47; Mismatches 154; Indels 112; Gaps 21;    |
| Db                    | 11          | EREDDGAVDRTDTRMVAFLGFLALYDSDGPGLALDPTWRAALPPNPINVDHARC 70  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Qy                    | 66          | EVGRVLAVNDPGRPFVFLIACVQLERVLTEAASAIIFERGPALSREERLYLTNYL 125  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Db                    | 71          | SVQVLAVNDARGPFVYVAGCQLEHVLEAASGDIIGRRGPPLTREERLYLTNYL 130  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Qy                    | 126         | PSVSLSTCRGDEV--PDRTLFAHYALCAIGRRGRTTYDTSLDAIAAPFRHLDPA 182   | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Db                    | 131         | PSVSLSTCRDSSAPVVGPN--I-FAHYALCAIGRRGRTTYDTMVEGAVAFRRIGLAS 188  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Db                    | 71          | REGVRREAAEELALAGRTWAPGEAHTLSTAVNMMLDRDSVLAERRRQAGIAGH 242  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Qy                    | 183         | TYLOASEKPKIWAESAPAPERYGKTAGMDS--PAASVAPQAVARQVAAS-- 297  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Db                    | 189         | TYLOQYLRDAVAEEAVSGERTWAPASARALTRTLATAVNMMMLDRDSVLAERRRQAGIAGH 248  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Db                    | 136         | TYLOASKEPKIWAESAPAPERYGKTAGMDS--PAASVAPQAVARQVAAS-- 297  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |
| Db                    | 249         | TYLOQASAKFGL--SDASP----RADRESAADQRHTRADGTQAPESAERRHEPSLRL 300  | Qy               | Query Match 40.3%; Score 1361; DB 2; Length 562;                      |



|                       |   |              |                              |
|-----------------------|---|--------------|------------------------------|
| Best Local Similarity | 40.3%   | Pred. No.    | 3.6e-41;                     |
| Matches               | 278;  | Conservative | 75;                          |
| Mismatches            | 212;  | Indels       | 125;                         |
| Gaps                  | 20;   |              |                              |
| /                     |   |              |                              |
| 18                    | AVPIVAGFGLALYDSDPGLALDDPTVRAALPPNPINVDARCEGRVLAVVNDP          | 77           | RA                           |
| 12                    | SLPIVAGTALYDNGDGGLTRETVAALPASRLPNNIDRNGCTVGEVLISVDDA          | 71           | RT                           |
|                       | RESEQUENCE FROM N.A.  |              | RT                           |
|                       | STRAIN=AB;  |              | RT                           |
|                       | Davison A.J.;   |              | RT                           |
|                       | Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.       |              | RT                           |
| 78                    | RGPFVGLIACVQLERVLETTAASAIAIFERGPALSREERPLYLITNLYLPSVSLSTKRRGD | 137          | RL                           |
| 72                    | RGPFGLIINCPQLGAVLAAAGPDEFGELSEGILSEQERLVLSPNLYLPSASLSSRLGP    | 131          | DR                           |
| 138                   | EVPPDRTLAVHVALCAIGRLGTVITYDLSLAAIAFPHLDATREGVREAAEFLAL        | 197          | DR                           |
| 132                   | DEBDETLFAHVSLCIVGRGTVITYDAPENAVAPFKRLSPSSREELLTNEAQSLI        | 191          | DR                           |
| 198                   | A-GRTWAPGVCEALTHLILLSTAVNNMLDRDWSLYAERRRGAIGTCHYLOASEKFKI     | 253          | KW                           |
| 192                   | GDAATWHLSSDTLRLVLLRWNLYLARRGEAGIETHYLASASFGITG                | 251          | Protase.                     |
| 254                   | -WGAESAPAPERGYKTKGAPGAMDTISPAASVAPAVARQVASS-                  | 298          | SEQUENCE                     |
| 252                   | CNKADFCGAEVLVDT-CGTYKSG-----EKVHGAPYSRVLGAKAFTSPNLPSSDN       | 302          | 646 AA;                      |
| 299                   | -SSSSFPAPADNPVSAAGAAPPGGSYLWIPQASLQVQHPP                      | 350          | Query Match                  |
| 303                   | DKGIGIGERICKHISAMASSNPOTLSAAGAFLVSGD-YIIVPAQYNLVQGQH-TSHPP    | 359          | 31 9%;                       |
| 351                   | LTAACGLPAACTVAYGHPGACPSPHYPAPPAPYPCGMLFACP-----SPLAQIAALV     | 402          | Score 1079.5;                |
| 360                   | INAGAPVTHAV-----PSQITPPAINSLMPMSYQAPPVSPHSANLEQITALV          | 411          | DB 2;                        |
| 403                   | GAIADROAGGLPAAAGDHIGRS-----AKRRRHEVEQFYDGRDDEPDR              | 448          | Length 646;                  |
| 412                   | GAIADRKRA--TKSOSDPHVQGSCSPPLSPQOERRYARKRHDWDATTRD-----DL      | 462          | Matches 278;                 |
| 449                   | DFPYDGEARPEPREVDSSRQAQSGPHETITALVGAVTSLOQELAHMRA-----RT       | 501          | Best Local Similarity 40.3%; |
| 463                   | EGIYKPGERSPRP--GERBAGR--PSTTIADLGMAVSLSQEVQLSQRRAIOTVTAQPQ    | 515          | Matches 75;                  |
| 502                   | HAPGYPPVPGPYHHPHADTEPQAOPPRYKAVYLPPPH-----TAPPPLSGAVPPP       | 556          | Conservative 75;             |
| 516                   | AAPAGLYKP-----PAVPPQI-SQYQIOPQAVASIAVAPQGI-----P              | 556          | Indels 125;                  |
| 557                   | SYPVVAUTCPAPPPLHOPSSPAHAAHPPPPGPTPPAASLQPEAERAGALYNASSA       | 616          | Gaps 20;                     |
| 557                   | SOPTQAVLAQFQVPAEGAKVYRAASTAQQAEOARAAPQFEAVTSAMVLPVTQPQAS      | 616          | RESULTS                      |
| 617                   | HWDVDTARA-----ANLVSMGSR                                       | 637          | 12                           |
| 617                   | SQTVDASASTGLEFGRDDADIFVSCMMSAR                                | 646          | PRELIMINARY;                 |
|                       | SEQUENCE FROM N.A.  |              | PRT;                         |
|                       | STRAIN=V92;   |              | 646 AA.                      |
|                       | Davis-Poynter N.J.; Nugent J.; Birch-Machin I.; Allen G.P.;   |              |                              |
|                       | Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.       |              |                              |
|                       | [1]   |              | SEQUENCE FROM N.A.           |
|                       | STRAIN=AB;  |              | SEQUENCE FROM N.A.           |
|                       | MEDLINE=12295566; PubMed=1318606;                             |              | SEQUENCE FROM N.A.           |
|                       | [2]   |              | SEQUENCE FROM N.A.           |
|                       | STRAIN=V92;   |              | SEQUENCE FROM N.A.           |
|                       | Davis-Poynter N.J.; Nugent J.; Birch-Machin I.; Allen G.P.;   |              | SEQUENCE FROM N.A.           |
|                       | Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.       |              | SEQUENCE FROM N.A.           |
|                       | [1]   |              | SEQUENCE FROM N.A.           |
|                       | STRAIN=AB;  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | AC  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE FROM N.A.           |
|                       | ID  |              | SEQUENCE FROM N.A.           |
|                       | AAS45919  |              | SEQUENCE                     |

|                       |  |   |      |  |
|-----------------------|--|---|------|--|
| DT                    | 03-MAR-2004  | (TREMBLrel. 27, Created)                | AC   | Q6X338;  |
| DT                    | 03-MAR-2004  | (TREMBLrel. 27, Last sequence update)   | DT   | 05-TUL-2004 (TREMBLrel. 27, Last sequence update)                      |
| DT                    | 03-MAR-2004  | (TREMBLrel. 27, Last annotation update) | DT   | 05-TUL-2004 (TREMBLrel. 27, Last annotation update)                    |
| DE                    | Capsid protein.  |   | DE   | UL26 capsid maturation serine protease.                                |
| OS                    | Equine herpesvirus 1.  |   | GN   | ORFnames-BHV5-31;  |
| OC                    | Viruses; dsDNA viruses, no RNA stage; Herpesviridae;                 |   | OS   | Bovine herpesvirus 5.  |
| OX                    | Alphaherpesvirinae; Varicellovirus.                                  |   | OC   | Viruses; dsDNA viruses, no RNA stage; Herpesviridae;                   |
| RN                    | NCBI_TaxID=10326;  |   | OC   | Alphaherpesvirinae.  |
| RP                    | [1]  |   | OX   | NCBI_TaxID=35344;  |
| RC                    | SEQUENCE FROM N.A.   |   | RN   | [1]  |
| RA                    | Davis-Poynter N.J., Nugent J., Birch-Machin I., Allen G.P.;          |   | RP   | SEQUENCE FROM N.A.   |
| RT                    | "Genome of bovine herpesvirus 5.";                                   |   | RC   | STRAIN=SV507/99;   |
| RL                    | Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.              |   | RT   | "Genome of bovine herpesvirus 5.";                                     |
| DR                    | EMBL;AY464052;AA65919.1; -.  |   | RL   | J. Virol. 77:10339-10347 (2003).                                       |
| SQ                    | SEQUENCE 646 AA; 68579 MW; 1C1C28C848116276 CRC64;                   |   | DR   | EMBL;AY226135; AAR86135.1; -.  |
| Query Match           | 31.9%; Score 1079.5; DB 2; Length 646;                               |   | DR   | GO; GO:0008233; F: Peptidase activity; IEA.                            |
| Best Local Similarity | 40.3%; Pred. No. 3-6e-41;  |   | DR   | InterPro; IPR01847; Peptidase S21.                                     |
| Matches               | 278; Conservative 75; Mismatches 212; Indels 125; Gaps 20;           |   | DR   | PFam; PF0716; Peptidase S21; I.  |
| Qy                    | 18 AVPIVAGFLALYDSDPGEIALDVTIALPPNPINVDIARCVGRVIAVNDP 77              |   | KW   | PRINTS; PRO0236; HSVCASTIDP40.   |
| Db                    | 12 SLPIVAGFLALYDMDGGLTRETTVALPPASRPPINIDERNGTCVGEVLISVDDA 71         |   | SQ   | SEQUENCE 619 AA; 63656 MW; 19275BBF269310B84 CRC64;                    |
| Qy                    | 7B RGPFFVGLIACVQLERVLETTAASAIAIFERFGPALSREERFLYLITNLYPSVSLSTKRGD 137 |   |      | Query Match 31.2%; Score 1056; DB 2; Length 619;                       |
| Db                    | 72 RGPFFVGLIACVQLERVLETTAASAIAIFERFGPALSREERFLYLITNLYPSVSLSTKRGD 131 |   |      | Best Local Similarity 41.4%; Pred. No. 4e-40;                          |
| Qy                    | 138 EVPPDRTLFAHVALCAIGRLGTIVYDTSLDAIAAPRFLHDPATREGYRREAEAILL 197     |   |      | Matches 270; Conservative 63; Mismatches 229; Indels 90; Gaps 19;      |
| Db                    | 132 DEEPDETLFAHVSCLCVGRVGTIVYDTPENAVAPFRLSPSSREELLTAREQSL 191        |   |      |  |
| Qy                    | 198 A-GRTWAPGVALTHTLSTAVNNNMRLDRWSLVAERRQAGTGHYQASENFKTI -- 253      |   | Qy   | 18 AVPIVAGFLALYDSDPGEIALDVTIALPPNPINVDIARCVGRVIAVNDP 77                |
| Db                    | 192 GDAATWHLSEDTLTVLLSTAVNNNLLRNRWNLVARRBEAGLEHTYQASASFGITNG 251     |   | Db   | 26 SMPVYGGTALYDGMDBEGLVTRARALPAPLPIINIDASACEGVAILAADD 85               |
| Qr                    | 254 -----WGAESAPAPERGKTKGTPGAMDTSPAASVPAQAVARQVASSS----- 298         |   | Qy   | 26 SMPVYGGTALYDGMDBEGLVTRARALPAPLPIINIDASACEGVAILAADD 85               |
| Db                    | 252 CNKADFCGABLVDT --CGYKSG-----EKVHGAPYSRVTGLKAFTSSSPNAPLSSD 302    |   | Qy   | 78 RGPFVFGVLIACVQLERVLETTAASAIAIFERFGPALSREERFLYLITNLYPSVSLSTKRGD 137  |
| Qy                    | 299 -----SSSSFPAPADNPVSAASGAPAPPGGDSYWIIPASHYNQLVGOSAPRHP 350        |   | Db   | 86 AGLFPGVYNCVNPOLADVLADVAPPAFFGADAPALAPRERFLVLSVNLPSVSLSSRLLAP 145    |
| Db                    | 303 DKGIGIGERTOKHISAMSNPOTLSAAGPLVSSD - YLVPAAQYNQLVVGHT - SHP 359   |   | Qy   | 138 EVPDRTLFAHVALCAIGRLGTIVYDTSLDAIAAPRFLHDATREGYRREAEAILL 197         |
| Qy                    | 351 LTACCLPACTVAYCHPGGACPSPHYPPPAHPYGMFLFAGP-----SPLRQIAALY 402      |   | Db   | 146 GEEADTGLFAHVALCVLGRVGTIVYDTPENAVAPFRLSPSSREELLTAREQSL 205          |
| Db                    | 360 INAGPAPVTHA-----PSQVTPPAYNLMPSMPSVPAQYWSVPHISNLQAQITAV 411       |   | Qy   | 198 AGRTWAPGVALTHTLSTAVNNNMRLDRWSLVAERRQAGTGHYQASEKFKINGAE 257         |
| Qy                    | 403 GAIADRQAGGLPAAGDGGIRGS-----AKRRRHEVEQEYDGRDDEPR 448              |   | Db   | 206 GDRAPVPREALRTLSTAVNNNMRLDRWSLVAERRQAGTGHYQASEKFKINGAE 265          |
| Db                    | 412 GALADRAKA---TKGSDPPIVIOG3QCSPSPLSQERRYKRRHDWDAITRD-----DL 462    |   | Qy   | 258 SAPAPERGYKTTGAGPMDTSPASVPAQAVARQVASSSSSSFPAPADNPVSAASCA 317        |
| Qy                    | 444 DFPYYPGEARPEPPRVDSRAARAOASGPBETITALVGAVTSLOELAHMRA-----RT 501    |   | Db   | 266 E-PERSRAQKAGLAGV-----CLALPVASGRARQPELPPA---PPAMPSAVHQASAS 314      |
| Db                    | 463 EGIYYPGEERSPR---GERRAGR-----PSTTIALMGMAVSSLQOEVQSRRAIQTVTAQQ 515 |   | Qy   | 318 PAPPBGDSYLMTPASHYNOVLYGQS-----APRH-BPLTAC-----GL 356               |
| Qy                    | 502 HAPYPGPVPGVPHPAHDTETPQPPRPAKAYTLPPH-----IAPPGPPPSGAVPP 556       |   | Db   | 315 PAHPLPA-GDVYTVPTAQYNQLVISQARGAAAAPPAPYFLPPAIAAPPMPGWTGA 373        |
| Db                    | 516 AAPGLYKEI-----PAVPQYI-----PAVPQYI-----PAVPQYI-----PAVPQYI 556    |   | Qy   | 357 PAAGTVAYGHPGAGPSPHYPPPAHPYGMFLFAGPSPLEAQTAALVGIAADR--OAGSL 414     |
| Qy                    | 557 SYPPVATPGPAPLHOPSPAHAHPPPPGTPPPAASLPOPEAPGAAGALVNASSA 616        |   | Db   | 374 AASCAAAPWHPGT-----FPPP-----GLESQTMALAGJADGRLRQAQGA 415             |
| Db                    | 557 SQPTQAVLAFQVPAEGAPSAKTYAATAPQQAEOQARAPLQVTPVQDAS 616             |   | Qy   | 415 --PAAGDGHIGRSKARRHVEQEYDGRDPEPDRDPEPYPGEARPEPPRVDSSRRAARQ 472      |
| Qy                    | 617 HVNDTARA-----ADLFLVSQMMGR 637                                    |   | Db   | 416 DGPYDGPGLDRRPAKRRYNEPLRSRGGD---DEAYYPOSEGGAPAEELPPQHRRPPSP 471     |
| Db                    | 617 SQTVDASATSTGLBFGRDDADIFVQMMGR 646                                |   | Qy   | 473 ASGPHEITITALVGSVLSQQLAEAHMRAHTHPYGPYPPVGPYTHPHADTETPAPQPRYPA 532   |
| Qy                    | 650 PPAASLQDPEAEGAEGALVNASSAHNVN---DTARAADLFVQMMGR 637               |   | Db   | 472 PQAPAHASRLASAVSLSIQQEYSQLRA -GYPGPAAFAAAQHHLP-----PPVPCPCLPQ 523   |
| Db                    | 650 PPAASLQDPEAEGAEGALVNASSAHNVN---DTARAADLFVQMMGR 637               |   | Qy   | 533 KAVYLLPPPHIAPPGLPSLGAQVPPST-----PPVAVTPGPAPPLHOPSPAHAHPPPPGGPT 589 |
| Qy                    | 650 PPAASLQDPEAEGAEGALVNASSAHNVN---DTARAADLFVQMMGR 637               |   | Db   | 524 QQ-QPQOYAAFPQWVAGPAPQAPQAPAPLPA-----A 569                          |
| Q6x238                | PRELIMINARY;   |   | PRT; | 619 AA.  |
| ID                    | Q6x238   |   |      |  |

RESULT 14  
Q6x238  
ID Q6x238

PRELIMINARY;

PRT;

619 AA.

550 PPAASLQDPEAEGAEGALVNASSAHNVN---DTARAADLFVQMMGR 637  
570 PPAASLQDPEAEGAEGALVNASSAHNVN---DTARAADLFVQMMGR 619

| RESULT 15   | Db | 570 PAPAAAGGPPBEPGAI - ATVDASAVAGLPLAQOQACDPADIFVQMMRHR 619 |
|---|----|---|
| AAR86135 PRELIMINARY; PRT; 619 AA.  |    |   |
| ID AAR86135;  |    |   |
| AC AAR86135;  |    |   |
| DT 02-MAR-2004 (TREMBLrel. 27, Created)   |    |   |
| DT 02-MAR-2004 (TREMBLrel. 27, Last sequence update)  |    |   |
| DT 02-MAR-2004 (TREMBLrel. 27, Last annotation update)  |    |   |
| DB U226 capsid maturation serine protease.  |    |   |
| GN BHV5-31.   |    |   |
| OS Bovine herpesvirus 5.  |    |   |
| OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;   |    |   |
| OC Alphaherpesvirinae.  |    |   |
| OX NCBI_TaxID:35244;  |    |   |
| RN [1]  |    |   |
| SEQUENCE FROM N A.  |    |   |
| RP STRAIN:SV507/99;   |    |   |
| RC MEDLINE=2250801; PubMed=12970418;  |    |   |
| RX Delhon G., Moraes M.P., Lu Z., Afonso C.L., Flores E.F., Weiblen R., Kutish G.F., Rock D.L.; "Genome of bovine herpesvirus 5.," J. Virol. 77:10339-10347 (2003). |    |   |
| RA RT DR EMBL; AY261359; AAR86135.1; -.   |    |   |
| KW Protease.  |    |   |
| SEQUENCE 619 AA; 63656 MW; 192758F26930BB4 CRC64;   |    |   |
| Query Match Score 31.2%; Score 1056; DB 2; Length 619;  |    |   |
| Best Local Similarity 41.4%; Pred. No. 4e-40;   |    |   |
| Matches 270; Conservative 63; Mismatches 229; Indels 90; Gaps 19;   |    |   |
| Qy 18 AVPIYYAGFLALYDSDGPGELALDPDTVRAALPENPLPINVDHARCEVRLAVNDP 77  |    |   |
| Db 26 SNEVYVGGYALYGMGDEGELVLTERRVARAIPPAAPLPIINIDHAACEGAVLAIADD 85  |    |   |
| Qy 78 RGPFVFGLIACVQLERVLETAASAAIFERGPALEERLLYLITINYLPSPSLSTKRRGD 137  |    |   |
| Db 86 AGLFFVGVNNCPQLADVLADYAHPAFFGADAPALAPRERFLYLVSNLYLPSSLSSRLAP 145   |    |   |
| Qy 138 EYPDPDTLFAYHALCATGRGLGTTVTDTSLODAIAFPFLDPATREGVREAAEELAL 197   |    |   |
| Db 146 GEEADGTLFAYHALCVLGRVGTIVTYDAPTEACVAPFRRLSPRAAIIADAEEAARAL 205  |    |   |
| Qy 198 AGRTWAPGEVALTHLILSTAVNMMLDRSLVAERRQAGTIAQHTYLOASEKFKIWGA 257   |    |   |
| Db 206 GDAWPVPREALARTLISAVNMVLKDWTVSRRREAGAHTYLOQAAVPLAGG 265   |    |   |
| Qy 258 SAPAPERGYKTGAPGAMDTSPAASVPAVQAVRARQAVASSSSSSFPAPADMNPVSGA 317  |    |   |
| Db 266 E-PERSRAQKGALLGGV----CIALFVSGARQPELPPA----PPAMSAVHOASAS 314  |    |   |
| Qy 318 PAPPNGDGSYLMIPASHYINQLYTGS ----- APRH_PPLTAC-----GL 356  |    |   |
| Db 315 PAHPLPA-GDYYVYVPTAQNYLIVSQARGAAAAAPPAAPYFTPPA AAAAPPMPGWYGA 373  |    |   |
| Qy 357 PAAGTVAYGHPGAGGSPHGPSPHPPPAHPPYFGMLFAGPSPLEQIAALYGAADR--QAGGL 414  |    |   |
| Db 374 AASGAAAATWHPGYG----FPPP-----GLESQ_MALAGAIADGRRLAQGA 415  |    |   |
| Qy 415 --PAAAGDGHGIRSAKRRREVEQOEPEYDCGRDDEPDRDFPPYYPPGEARPEPRPVDSRRAARQ 472   |    |   |
| Db 416 DGGYDGPIDRPPKERRRNWPLRSRGGD---DEAYYFEGEGAPAEELPQHHHRPPSP 471   |    |   |
| Qy 473 ASGPHEITTAIYCAVTSIQLQELAHMRAARTHAPYGPYPVGPYTHPHADTETPAQPPRYPA 532  |    |   |
| Db 472 PQAPAHALSRLASAVSIIQQEVSQRLA--GYFYGPAFAAAQHHLP----PAPVCLPQ 523  |    |   |
| Qy 533 KAVYLLPPPHIAPPGPPLSGLQAVPPS-----PPVAVTPGPAPLHQPSPAHAHBPPPPBPGT 589   |    |   |
| Db 524 QQ-QPQQWYAPQPVVAGPAPAPAQALAPAPQAPAPLATPA-----A 569   |    |   |
| Qy 590 PPPAASLUPQPEAPGAEAGLYNASSAAHNVN---DTARAADLFVQSMGMSR 637  |    |   |

Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 10, 2005, 19:08:55 ; Search time 40 Seconds (without alignments)

1056.114 Million cell updates/sec

Title: US-10-623-429-9

Perfect score: 3385

Sequence: 1 MASAMERLLEAPLPPDRAVP.....VNVDTARAADLFVSSQMMGSR 637

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0\*

Maximum Match 100\*

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cn2\_6\_ptodata/1/iaa/5A COMB.pep:\*

2: /cn2\_6\_ptodata/1/iaa/5B COMB.pep:\*

3: /cn2\_6\_ptodata/1/iaa/6A COMB.pep:\*

4: /cn2\_6\_ptodata/1/iaa/6B COMB.pep:\*

5: /cn2\_6\_ptodata/1/iaa/ACTUS COMB.pep:\*

6: /cn2\_6\_ptodata/1/iaa/backfles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match | Length | DB               | ID               | Description       |                    |                    |
|------------|-------|-------|-------|--------|------------------|------------------|-------------------|--------------------|--------------------|
| 1          | 23.3  | 68.9  | 635   | 4      | US-08-176-320-2  |                  | Sequence 2, Appli |                    |                    |
| 2          | 23.3  | 68.6  | 635   | 1      | US-07-832-855-2  |                  | Sequence 2, Appli |                    |                    |
| 3          | 14.2  | 42.3  | 350   | 1      | US-08-458-067-2  |                  | Sequence 2, Appli |                    |                    |
| 4          | 14.2  | 42.3  | 350   | 5      | PCT-US96-07795-2 |                  | Sequence 2, Appli |                    |                    |
| 5          | 14.2  | 42.3  | 350   | 5      | PCT-US96-07796-2 |                  | Sequence 2, Appli |                    |                    |
| 6          | 12.62 | 37.3  | 306   | 1      | US-08-279-754-2  |                  | Sequence 2, Appli |                    |                    |
| 7          | 12.62 | 37.3  | 305   | 6      | PCT-US95-09052-2 |                  | Sequence 2, Appli |                    |                    |
| 8          | 12.52 | 37.0  | 247   | 3      | US-03-167-434-4  |                  | Sequence 4, Appli |                    |                    |
| 9          | 12.52 | 37.0  | 247   | 3      | US-08-853-755-4  |                  | Sequence 4, Appli |                    |                    |
| 10         | 11.26 | 33.3  | 247   | 3      | US-09-167-434-10 |                  | Sequence 3, Appli |                    |                    |
| 11         | 11.26 | 33.3  | 247   | 3      | US-08-853-755-3  |                  | Sequence 3, Appli |                    |                    |
| 12         | 7.87  | 23.2  | 663   | 1      | US-07-912-015-2  |                  | Sequence 2, Appli |                    |                    |
| 13         | 5.90  | 5     | 17.4  | 243    | 3                | US-09-167-434-9  |                   | Sequence 9, Appli  |                    |
| 14         | 5.90  | 5     | 17.4  | 243    | 3                | US-08-853-755-9  |                   | Sequence 9, Appli  |                    |
| 15         | 5.90  | 5     | 17.4  | 255    | 3                | US-09-167-434-10 |                   | Sequence 10, Appli |                    |
| 16         | 5.90  | 5     | 17.4  | 255    | 3                | US-08-853-755-10 |                   | Sequence 10, Appli |                    |
| 17         | 5.86  | 5     | 17.3  | 236    | 3                | US-03-167-434-5  |                   | Sequence 5, Appli  |                    |
| 18         | 5.86  | 5     | 17.3  | 236    | 3                | US-08-853-755-5  |                   | Sequence 5, Appli  |                    |
| 19         | 5.86  | 5     | 17.3  | 242    | 3                | US-09-167-434-8  |                   | Sequence 8, Appli  |                    |
| 20         | 5.86  | 5     | 17.3  | 242    | 3                | US-08-853-755-8  |                   | Sequence 8, Appli  |                    |
| 21         | 5.86  | 5     | 17.3  | 246    | 3                | US-03-167-434-11 |                   | Sequence 11, Appli |                    |
| 22         | 5.86  | 5     | 17.3  | 246    | 3                | US-08-853-755-11 |                   | Sequence 11, Appli |                    |
| 23         | 5.86  | 5     | 17.3  | 254    | 3                | US-03-167-434-7  |                   | Sequence 7, Appli  |                    |
| 24         | 5.86  | 5     | 17.3  | 254    | 3                | US-08-853-755-7  |                   | Sequence 7, Appli  |                    |
| 25         | 3.29  | 5     | 9.7   | 609    | 3                | US-08-251-288A-2 |                   | Sequence 2, Appli  |                    |
| 26         | 3.29  | 5     | 9.7   | 609    | 3                | US-09-298-819A-2 |                   | Sequence 2, Appli  |                    |
| 27         | 3.29  | 5     | 9.7   | 609    | 4                | US-09-586-563C-2 |                   | Sequence 2, Appli  |                    |
|            |       |       |       | 29     | 329.5            | 9.7              | 609               | 4                  | US-09-586-562C-2   |
|            |       |       |       | 30     | 268.5            | 7.9              | 802               | 4                  | US-09-823-240A-2   |
|            |       |       |       | 31     | 255.5            | 7.5              | 720               | 3                  | US-09-219-849-4    |
|            |       |       |       | 32     | 255.5            | 7.5              | 777               | 1                  | US-08-642-255-53   |
|            |       |       |       | 33     | 255              | 7.5              | 504               | 3                  | US-09-219-849-3    |
|            |       |       |       | 34     | 255              | 7.5              | 561               | 1                  | US-08-642-255-52   |
|            |       |       |       | 35     | 253.5            | 7.5              | 1184              | 4                  | US-09-266-25D-18   |
|            |       |       |       | 36     | 253.5            | 7.5              | 1185              | 3                  | US-09-041-886-23   |
|            |       |       |       | 37     | 253.5            | 7.5              | 1185              | 4                  | US-09-538-092-1209 |
|            |       |       |       | 38     | 248.5            | 7.3              | 1274              | 3                  | US-09-095-443-2    |
|            |       |       |       | 39     | 242              | 7.1              | 408               | 1                  | US-07-609-716-65   |
|            |       |       |       | 40     | 242              | 7.1              | 408               | 3                  | US-08-475-411A-65  |
|            |       |       |       | 41     | 242              | 7.1              | 408               | 3                  | US-08-478-029A-55  |
|            |       |       |       | 42     | 238.5            | 7.0              | 330               | 1                  | US-08-642-255-32   |
|            |       |       |       | 43     | 236              | 7.0              | 960               | 3                  | US-09-219-849-5    |
|            |       |       |       | 44     | 234.5            | 6.9              | 230               | 3                  | US-09-064-703-7    |
|            |       |       |       | 45     | 229.5            | 6.8              | 424               | 4                  | US-09-538-092-1338 |

ALIGNMENTS

RESULT 1

US-08-176-320-2

; Sequence 2, Application US/08176320

; Patent No. 641074

; GENERAL INFORMATION:

; APPLICANT: Roizman, Bernard

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE PREPARATION AND USE OF A HERPES PROTEASE

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Alice O. Martin

; STREET: 321 No. 6410704th Clark Street, Suite 800

; CITY: Chicago

; STATE: Ill

; COUNTRY: USA

; ZIP: 60610

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/176,320

; FILING DATE:

; CLASSIFICATION: 435

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 07/705,814

; FILING DATE: 24-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Cooley, Ronald B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (312)744-0090

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 635 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-176-320-2

Query Match 68.9%; Score 2333; DB 4; Length 635;

Best Local Similarity 74.0%; Prd. No. 1..9e-160;

Matches 477; Conservative 30; Mismatches 120; Indels 18; Gaps 7;

Qy 1 MASAEMRRLLEAPLPPDRAVPITVAGFLALYDSSDGPGEIALDPPDTVRAALPPNPLPINVD 60

Db 1 MAAADPGRDRMEEELPDRAPIVAGFLALYDSSDGPGEIALDPPDTVRAALPPNPLPINVD 60

Qy 61 HRARCEYGRVLAVYNDPRGPFVYGLIACVQLERVLTAASAAIFRRGPALSREERULYL 120



ZIP: 07065-0907  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/458,067  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Giesser, Joanne M.  
 REGISTRATION NUMBER: 32,838  
 REFERENCE/DOCKET NUMBER: 19457  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (908) 594-3046  
 TELEFAX: (908) 594-4720  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 350 amino acids  
 TYPE: amino acid  
 TOPOLogy: linear  
 MOLECULE TYPE: protein  
 US-08-458,067-2

Query Match 42.3%; Score 1432; DB 1; Length 350;  
 Best Local Similarity 80.5%; Pred. No. 9.9e-96;  
 Matches 285; Conservative 13; Mismatches 46; Indels 10; Gaps 2;

Qy 1 MASADMREPLEAPLDPDRAPVPIVAGFLAYDSDGPGEALDPDTVRAALPPNPPLPINVD 60  
 Db 1 MAADAPGDRMEEPLPDRAPVPIVAGFLAYDSDGPGEALDPDTVRAALPPNPPLPINVD 60

Qy 61 HRARCEVGRVLAIVNDPDRPFFVGLIACVQLERVLETASAALFERRGPALEERLLYL 120  
 Db 61 HRARCEVGRVLAIVNDPDRPFFVGLIACVQLERVLETASAALFERRGPALEERLLYL 120

Qy 121 ITNYLPSVSLSTKRGDEVPPDRTLFAYVALCAIGRRLGTTVYDTSDDAAIAAPRHLDP 180  
 Db 121 ITNYLPSVSLATKRGGEAHPDRTLFAYVALCAIGRRLGTTVYDTSDDAAIAAPRHLSP 180

Qy 181 ATREGVRRBAAEAEELLAGRTWAPGEALTHTLISTAVNMMLDRWSLVAERRQAGIA 240  
 Db 181 ASREGARRLAAEAEELSGRTWAPGEALTHTLISTAVNMMLDRWSLVAERRQAGIA 240

Qy 241 GHTYLOASEKFWKIGAESAPAPERYGKTYGKAPGMDTSPASVPAHQ---VAVRAVOVAS 296  
 Db 241 GHTYLOASEKFWKIGAESAPAPERYGKTYGKAPGMDTSPASVPAHQ---VAVRAVOVAS 296

Qy 297 SSSSSFPAPADMNPPVSAASGAPAPPGGDSYLMIPASHYNOLVTGOSAPRHP 350  
 Db 301 P-----PVLPKMNPPVTSGTPAPPGDGSYLMIPASHYNOLVTGOSAPRHP 348

Qy 297 SSSSSFPAPADMNPPVSAASGAPAPPGGDSYLMIPASHYNOLVTGOSAPRHP 350  
 Db 301 P-----PVLPKMNPPVTSGTPAPPGDGSYLMIPASHYNOLVTGOSAPRHP 348

RESULT 5  
 PCT-US96-07796-2

RESULT 4  
 PCT-US96-07795-2  
 ; Sequence 2, Application PC/TUSS9607795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MERCK & CO., INC.  
 ; APPLICANT: Register, Robert B.  
 ; APPLICANT: Shafer, Jules A.  
 ; TITLE OF INVENTION: HERPES SIMPLEX TYPE 1 PROTEASE MUTANTS  
 ; TITLE OF INVENTION: HERPES SIMPLEX TYPE 1 PROTEASE MUTANTS  
 ; NUMBER OF SEQUENCES: 40  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Ms. Joanne M. Giesser  
 ; STREET: 126 East Lincoln Avenue, P.O. Box 2000-0907  
 ; CITY: Rahway  
 ; STATE: New Jersey  
 ; COUNTRY: US  
 ; ZIP: 07065-0907  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US96/07796  
 FILING DATE: 07/09/1996  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Giesser, Joanne M.  
 REGISTRATION NUMBER: 32,838  
 TELECOMMUNICATION INFORMATION:  
 REFEERENCE/DOCKET NUMBER: 19233  
 TELEPHONE: (908) 594-47046  
 TELEFAX: (908) 594-4720  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 306 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-07796-07796-2

Query Match 42.3%; Score 1432; DB 5; Length 350;  
 Best Local Similarity 90.5%; Freq. No. 9; e-96%; Indels 10; Gaps 2;  
 Matches 285; Conservative 13; Mismatches 46; Indels 1; Gaps 1;

Qy 1 MASAEMBERLEAPLDRAPIVAGFLALYDSDGSGFLALDPDTVRAALPBNPLPINVD 60  
 Db 1 MADAAPGDRMEBPLPDRAPIVAGFLALYDSDGSGFLALDPDTVRAALPBNPLPINVD 60  
 Db 1 HRAGECYGRVLAVVNDPRGPFFVGLIACVQLERVELTAAASAIIFERGPALSREBLLYL 120  
 Db 61 HRAGECYGRVLAVVNDPRGPFFVGLIACVQLERVELTAAASAIIFERGPALSREBLLYL 120  
 Qy 61 HARCEVGRVLAVVNDPRGPFFVGLIACVQLERVELTAAASAIIFERGPALSREBLLYL 120  
 Db 61 ITNYLPSVSLSTKRGDDEVPPDRTLPFAHVALCAIGRLGTVYDTSLDAAIAFPHLDP 180  
 Qy 121 ITNYLPSVSLSTKRGDDEVPPDRTLPFAHVALCAIGRLGTVYDTSLDAAIAFPHLDP 180  
 Db 121 ITNYLPSVSLATKRLGSEAHAPDRTLPFAHVALCAIGRLGTVYDTSLDAAIAFPHLSP 180  
 Qy 181 ATREGYRREAAELALAGRTWAPGVEALTHTLSTAVNNNMRLDRWSLVAERRGAGIA 240  
 Db 181 ASREGARRLAABELALSGRTWAPGVEALTHTLSTAVNNNMRLDRWSLVAERRGAGIA 240  
 Qy 241 GHTYLOSEKPKWIGASSAPAPERGKTAGPAMDTSPAASVPAQ ---VAVRAQVAS 296  
 Db 241 GHTYLOSEKPKWIGASSAPAPERGKTAGPAMDTSPAASVPAQ ---VAVRAQVAS 296  
 Qy 297 SSSSSSPPADMNPPVSAAGAPAPPPGDSYLWIPASHKNLVGTGOSAPRHP 350  
 Db 301 P-----PVLPMMNPVPTSGTAPAPPGDSYLWIPASHNOLVGAHPAQQPQP 348

RESULT 6  
 US-08-279-754-2  
 Sequence 2, Application US/08279754  
 Patent No. 5466470  
 GENERAL INFORMATION:  
 ADDRESSEE: JOANNE M. GIESSER  
 STREET: P.O. BOX 2000, 126 E. LINCOLN AVE.  
 CITY: RAHWAY  
 STATE: NJ  
 COUNTRY: USA  
 ZIP: 07065  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/279,754  
 FILING DATE:

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: GIESSER, JOANNE M.  
 REGISTRATION NUMBER: 32,838  
 TELECOMMUNICATION INFORMATION:  
 REFEERENCE/DOCKET NUMBER: 19233  
 TELEPHONE: (908) 594-47046  
 TELEFAX: (908) 594-4720  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 306 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-279-754-2

Query Match 37.3%; Score 1262; DB 1; Length 306;  
 Best Local Similarity 84.0%; Freq. No. 1; e-83%; Indels 9; Gaps 1;  
 Matches 252; Conservative 11; Mismatches 33; Indels 9; Gaps 1;

Qy 1 MASAEMBERLEAPLDRAPIVAGFLALYDSDGSGFLALDPDTVRAALPBNPLPINVD 60  
 Db 1 MADAAPGDRMEBPLPDRAPIVAGFLALYDSDGSGFLALDPDTVRAALPBNPLPINVD 60  
 Db 1 HRAGECYGRVLAVVNDPRGPFFVGLIACVQLERVELTAAASAIIFERGPALSREBLLYL 120  
 Db 61 HRAGECYGRVLAVVNDPRGPFFVGLIACVQLERVELTAAASAIIFERGPALSREBLLYL 120  
 Qy 121 ITNYLPSVSLSTKRGDDEVPPDRTLPFAHVALCAIGRLGTVYDTSLDAAIAFPHLDP 180  
 Db 121 ITNYLPSVSLATKRLGSEAHAPDRTLPFAHVALCAIGRLGTVYDTSLDAAIAFPHLSP 180  
 Qy 181 ATREGYRREAAELALAGRTWAPGVEALTHTLSTAVNNNMRLDRWSLVAERRGAGIA 240  
 Db 181 ASREGARRLAABELALSGRTWAPGVEALTHTLSTAVNNNMRLDRWSLVAERRGAGIA 240  
 Qy 241 GHTYLOSEKPKWIGASSAPAPERGKTAGPAMDTSPAASVPAQ ---VAVRAQVAS 296  
 Db 241 GHTYLOSEKPKWIGASSAPAPERGKTAGPAMDTSPAASVPAQ ---VAVRAQVAS 296

RESULT 7  
 PCT-0795-09052-2  
 Sequence 2, Application PC/TUS9509052  
 GENERAL INFORMATION:  
 APPLICANT: DARKE, PAUL L.  
 APPLICANT: HALL, DAWN L.  
 APPLICANT: KUO, LAWRENCE C.  
 TITLE OF INVENTION: PURIFIED HERPES SIMPLEX VIRAL PROTEASE  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: 126 E. LINCOLN AVE.  
 CITY: RAHWAY  
 STATE: NJ  
 COUNTRY: USA  
 ZIP: 07065  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/09052  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: GIESSER, JOANNE M.  
 REGISTRATION NUMBER: 32,838  
 TELECOMMUNICATION INFORMATION:  
 REFEERENCE/DOCKET NUMBER: 19233

TELEPHONE: (908) 594-3046  
 TELEFAX: (908) 594-4720  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 306 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO

PCN-US95-09052-2

Query Match 37.3%; Score 1262; DB 5; Length 306;  
 Best Local Similarity 84.0%; Pred. No. 1.6e-83;  
 Matches 252; Conservative 11; Mismatches 33; Indels 4; Gaps 1;

Qy 1 MASAEMLERLEAPLPDRAWPIVYAGFLALYDSDGPGLALDPPTVRAALPPNPLPINVD 60  
 Db 1 MAADAPGDRMEEPEPDPDRAWPIVYAGFLALYDSDGPGLALDPPTVRAALPPNPLPINVD 60

Qy 61 HRACEVGRVLAIVNDPQGPFVGLIACVQLERVLETAASAAIFERRSPALSREERLLY 120  
 Db 61 HRAGCEVGRVLAIVNDPQGPFVGLIACVQLERVLETAASAAIFERRSPALSREERLLY 120

Qy 121 ITNYLPVSLSLKRGDEVPPDRTLFATVALCATGRRGTTVYDTSLDAATAPFRHLDP 180  
 Qy 121 ITNYLPVSLSLKRGDEVPPDRTLFATVALCATGRRGTTVYDTGIDAAATAPFRHLDP 180

Db 121 ITNYLPVSLSLKRGDEVPPDRTLFATVALCATGRRGTTVYDTGIDAAATAPFRHLDP 180

Qy 181 ATREGVRRPAAEELALLGRTWAPGVVEALTHILSTAVNNMLDRNSLVAERRQGIA 240  
 Db 181 ASNREGARRLAAEELALSQRTWAPGVVEALTHILSTAVNNMLDRNSLVAERRQGIA 240

Qy 241 GHYIQLQASSEKFKIWGAESAPAPRGYKIGAPGAMDTSPAASYAPQO --- -VAVRAROVA 296  
 Db 241 GHYIQLQASSEKFKIWGAESAPAPRGYKIGAPGAMDTSPAASYAPQO --- -VAVRAROVA 300

RESULT 8  
 US-09-167-434-4  
 Sequence 4, Application US/09167434  
 Patent No. 600833  
 GENERAL INFORMATION  
 APPLICANT: Abdel Meguid, Sherin  
 APPLICANT: Culp, Jeffrey  
 APPLICANT: Debouck, Christine  
 APPLICANT: Hoog, Susan S.  
 APPLICANT: Janson, Cheryl A.  
 APPLICANT: Smith, Ward W.  
 TITLE OF INVENTION: No. 6008033el Proteases, Compositions Capable of  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SmithKline Beecham Corporation  
 STREET: 709 Swedeland Road - P.O. Box 1539  
 CITY: King of Prussia  
 STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19406-2799  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.3.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/167,434  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/853,755  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/022,470

RESULT 9  
 US-08-853-755-4  
 Sequence 4, Application US/08853755  
 Patent No. 6083711  
 GENERAL INFORMATION:  
 APPLICANT: Abdel-Meguid, Sherin  
 APPLICANT: Oiu, Xiayang  
 APPLICANT: Culp, Jeffrey  
 APPLICANT: Debouck, Christine  
 APPLICANT: Janson, Cheryl A.  
 APPLICANT: Hoog, Susan S.  
 APPLICANT: Smith, Ward W.  
 TITLE OF INVENTION: No. 6083711el Proteases, Compositions Capable of  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SmithKline Beecham Corporation  
 STREET: 709 Swedeland Road - P.O. Box 1539  
 CITY: King of Prussia

STATE: Pennsylvania  
 COUNTRY: USA  
 ZIP: 19406-2799  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/853,755  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/018,616  
 FILING DATE: 15-MAY-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/022,470  
 FILING DATE: 26-JUL-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/024,416  
 FILING DATE: 21-AUG-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/030,901  
 FILING DATE: 14-NOV-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/035,973  
 FILING DATE: 21-JUN-1997  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/039,191  
 FILING DATE: 27-FEB-1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Dinner, Dara L.  
 REGISTRATION NUMBER: 33,680  
 REFERENCE DOCKET NUMBER: P50472-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 610-270-5022  
 TELEFAX: 610-270-5090  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 247 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-853-755-4

Qy 1 MASAEMERLEAPLDRAVPIVYAGFLALYDSDGPGFLALDDPTVRLAPLPENPFLINVD 60  
 Db 1 MASAEMERLEAPLDRAVPIVYAGFLALYDSDGPGFLALDDPTVRLAPLPENPFLINVN 60  
 Qy 61 HARCEVGRVLAVNDRGPFFPGFLACVQLERVLEAAAIIFERGPALSREERFLYL 120  
 Db 61 HARCEVGRVLAVNDRGPFFPGFLACVQLERVLEAAAIIFERGPALSREERFLYL 120  
 Qy 121 ITNLPVSLSLSTKRGDEVPPDRTLFAHVALCAIGRLGTVYDTSLDAATAPFRHLDP 180  
 Db 121 ITNLPVSLSLSTKRGDEVPPDRTLFAHVALCAIGRLGTVYDTSLDAATAPFRHLDP 180  
 Qy 181 ATPREGVREAAEALALAGPTWARGFLAATLTHLITSTAVNNMMRLDWSLVAERRQAGIA 240  
 Db 181 ATREGVREAAEALALAGRTWAPGEALTHLITSTAVNNMMRLDWSLVAERRQAGIA 240  
 Qy 241 GHLYLQA 247  
 Db 241 GHLYLQA 247

Qy 1 MASAEMERLEAPLDRAVPIVYAGFLALYDSDGPGFLALDDPTVRLAPLPENPFLINVD 60  
 Db 1 MADAOPDRMELPLPDRAVPIVAGFLALDDPTVRLAPLPENPFLINVN 60  
 Qy 61 HARCEVGRVLAVNDRGPFFPGFLACVQLERVLEAAAIIFERGPALSREERFLYL 120

; Sequence 3, Application US/09167434  
; Patent No. 6008033  
; GENERAL INFORMATION:  
; APPLICANT: Abel-Meguid, Sherin  
; APPLICANT: Qiu, Xiaoyang  
; APPLICANT: Culp, Jeffrey  
; APPLICANT: Debouck, Christine  
; APPLICANT: Janson, Cheryl A.  
; APPLICANT: Hoog, Susan S.  
; APPLICANT: Smith, Ward W.  
; TITLE OF INVENTION: No. 6008033el Proteases, Compositions Capable of  
; TITLE OF INVENTION: Binding to Said Site, and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Smithline Beecham Corporation  
; STREET: 709 Swedeland Road - P.O. Box 1539  
; CITY: King of Prussia  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19406-2799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/167,434  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/853,755  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/022,470  
; FILING DATE: 26-JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/024,416  
; FILING DATE: 21-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/030,901  
; FILING DATE: 14-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/035,973  
; FILING DATE: 21-JAN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,191  
; FILING DATE: 27-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dinner, Dara L.  
; REGISTRATION NUMBER: 33,680  
; REFERENCE DOCKET NUMBER: P50472-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-5022  
; TELEFAX: 610-270-5090  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 247 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-167-434-3

Query Match 33.3%; Score 1126; DB 3; Length 247;  
 Best Local Similarity 89.4%; Pred. No. 8-74; Indels 9; Gaps 0;  
 Matches 246; Conservative 1; Mismatches 0;

Qy 1 MASAEMERLEAPLDRAVPIVYAGFLALYDSDGPGFLALDDPTVRLAPLPENPFLINVD 60  
 Db 1 MASAEMERLEAPLDRAVPIVYAGFLALYDSDGPGFLALDDPTVRLAPLPENPFLINVN 60  
 Qy 61 HARCEVGRVLAVNDRGPFFPGFLACVQLERVLEAAAIIFERGPALSREERFLYL 120  
 Db 61 HARCEVGRVLAVNDRGPFFPGFLACVQLERVLEAAAIIFERGPALSREERFLYL 120  
 Qy 121 ITNLPVSLSLSTKRGDEVPPDRTLFAHVALCAIGRLGTVYDTSLDAATAPFRHLDP 180  
 Db 121 ITNLPVSLSLSTKRGDEVPPDRTLFAHVALCAIGRLGTVYDTSLDAATAPFRHLDP 180  
 Qy 181 ATPREGVREAAEALALAGPTWARGFLAATLTHLITSTAVNNMMRLDWSLVAERRQAGIA 240  
 Db 181 ATREGVREAAEALALAGRTWAPGEALTHLITSTAVNNMMRLDWSLVAERRQAGIA 240  
 Qy 241 GHLYLQA 247  
 Db 241 GHLYLQA 247

Query Match 33.3%; Score 1126; DB 3; Length 247;  
 Best Local Similarity 89.4%; Pred. No. 8-74; Indels 9; Gaps 0;  
 Matches 222; Conservative 9; Mismatches 16;

Qy 1 MASAEMERLEAPLDRAVPIVYAGFLALYDSDGPGFLALDDPTVRLAPLPENPFLINVD 60  
 Db 1 MADAOPDRMELPLPDRAVPIVAGFLALDDPTVRLAPLPENPFLINVN 60  
 Qy 61 HARCEVGRVLAVNDRGPFFPGFLACVQLERVLEAAAIIFERGPALSREERFLYL 120

Db 61 HRAGCEVGRVLAVVDDPRGPFFVGLIACVQLERVLETAASAAIFERRGPPLSREERLLYL 120 ; LENGTH: 247 amino acids  
 Qy 121 ITNYLPSVLSTKRGDEYPPDRTLFAHYVALCAIGRRGTTVYDTSLDAATAPFRHIDP 180 ; TYPE: amino acid  
 Db 121 ITNYLPSVLSTKRGDEYPPDRTLFAHYVALCAIGRRGTTVYDTGLDAATAPFRHIDP 180 ; STRANDBEDNESS: ;  
 ; TOPOLOGY: linear  
 ; MOLCODE: protein  
 US-08-853-755-3

Qy 181 ATREGVRRRAAEALAGRTWAQVEALTHLILSTAVNMMLRDRWSVAERRQAGIA 240 ; Query Match 33.3%; Score 1126; DB 3; Length 247;  
 Db 181 ASREGRRLAAEALAGRTWAQVEALTHLILSTAVNMMLRDRWSVAERRQAGIA 240 ; Best Local Similarity 89.9%; Prod. No. 8e-74;  
 ; Missmatches 16; Indels 0; Gaps 0;

Qy 241 GHTYLOA 247 ; Matches 222; Conservative 9;

Db 241 GHTYLOA 247 ; 1 MASABMRERLEAPLIDPRAPIVYAGFLALYDSCDPGEIADPDTVAALPPENPLPINVD 60  
 ; 1 MAADAPGDMEEPPLIDPRAPIVYAGFLALYDSCDPGEIADPDTVAALPPDNPLPINN 60

---

RESULT 11 ; Query 61 HRACCEVERVLAVYNDPQGPFFVGLIACVQLERVLETAASAAIFERRGPPLSREERLLYL 120  
 US-08-853-755-3 ; Database 61 HRACCEVERVLAVYNDPQGPFFVGLIACVQLERVLETAASAAIFERRGPPLSREERLLYL 120  
 ; Sequence 3, Application US/0883755  
 ; Patent No. 6083711

; GENERAL INFORMATION:  
 ; APPLICANT: Abdel-Meguid, Sherin  
 ; APPLICANT: Qiu, Xiayang  
 ; APPLICANT: Culp, Jeffrey  
 ; APPLICANT: Debouck, Christine  
 ; APPLICANT: Janson, Cheryl A.  
 ; APPLICANT: Hoog, Ward W.  
 ; APPLICANT: Smith, Susan S.

; TITLE OF INVENTION: No. 6033711el Proteases, Compositions Capable of  
 ; TITLE OF INVENTION: Binding to Said Site, and Methods of Use Thereof  
 ; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Smithkline Beecham Corporation  
 ; STREET: 709 Swedeland Road - P.O. Box 1539  
 ; CITY: King of Prussia  
 ; STATE: Pennsylvania  
 ; COUNTRY: USA  
 ; ZIP: 19406-2799

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.3.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/853,755  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/018,616  
 ; FILING DATE: 15-MAY-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/022,470  
 ; FILING DATE: 26-JUL-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/024,416  
 ; FILING DATE: 21-AUG-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/030,901  
 ; FILING DATE: 14-NOV-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/035,973  
 ; FILING DATE: 21-JAN-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/039,191  
 ; FILING DATE: 27-FEB-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Dinner, Dara L.  
 ; REGISTRATION NUMBER: 33,680  
 ; REFERENCE/DOCKET NUMBER: P50472-1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-270-5022  
 ; TELEFAX: 610-270-5030  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 663 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear

RESULT 12 ; Application US/07912015  
 US-07-912-015-2 ; Sequence 2, Application US/07912015  
 ; Patent No. 5283191  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Morgan, Robin Wilson  
 ; APPLICANT: Willimse, Martha Jacoba  
 ; APPLICANT: Claessens, Johannes Antonius Joseph  
 ; APPLICANT: Sondermeijer, Paulus Jacobus Antonius  
 ; TITLE OF INVENTION: Marek's Disease Virus vaccine  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Organon Technika Corporation  
 ; STREET: 1330-A Piccard Drive  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.24  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/912,015  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/615,211  
 ; FILING DATE: 16-NOV-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Doma Bobrowicz  
 ; REGISTRATION NUMBER: 32,196  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 255-5200  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 663 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear

MOLECULE TYPE: protein  
 US-07-912-015-2

Query Match 23.2%; Score 787; DB 1; Length 663;  
 Best Local Similarity 31.6%; Pred. No. 7.5e-49;  
 Matches 229; Conservative 77; Mismatches 235; Indels 184; Gaps 26;

Qy 21 IYVAGFIALYDSDGPGLALDDPTVRAALPPENPLPNTDHRARCEYGRVIAVNVNDRGP 80  
 Db 9 VVAGVIALYLGADSEEDNIDKDIRAIIPTAPLPLNIDERRDCTVGAIALIDDEHGL 68

Qy 81 PFGVGLIACVQLERVLETAASAAIFERRGPAISREERLLYLITNLPVSLSLTKRCDEPV 140  
 Db 69 FFLGKINCPVWVTRLETAASQBTIFSEL-DLKPKDKLULYTITNLPVSLSRRLAPGET 127

Qy 141 PDRTLFIAHVALCAIGRLGILTVYDTSLDAATAPFRHLDPATREGYRREAAEALAGR 200  
 Db 128 ADETFIAHVALCLIGRIGITVYDLTPEATEPPFKLSPNSKATLSSQGKETERILGEM 187

Qy 201 TWAPGVVEALTHLTLSTAVNNNMMLRDWSLVAERRRQAGIAGHTYLOASEKEPKWGAESAP 260  
 Db 188 WYPSKQNAIKTALLGTAVNNMMLRDWQIISERRMAGITQKYLQASS-----236

Qy 261 APERGYKTGAPGAMDTPSAYSPAP-----  
 Db 237 -----FTALTDMS-TSNNVSYTHPICEANAFGNIQKDEENQCVISPAOTSETLNAGVL 289

Qy 299 SSSSFPAADMNPSAS-----  
 Db 290 GNDFFHLPHSDPASTSDQTNLQSLIEPSMTQSSRPGD-DFIWPKISYNQLYSRNNS 348

Qy 341 -----TGOSAPRHPPLTACGHPAAGTVAHYGHPGAGPSPHYHPP-----PAH 381  
 Db 349 QPTNIPDIAITSNQNPPTIPALMNTISGCHSISPSGTAQYQG-----YPTPGTHNSLPLG 404

Qy 382 P-----  
 Db 405 PYNQMGFGQYGPQVYPLSICOSPLEAKTALLECMTKERPVDEEHRGDDMHTTREER 427

Qy 428 KRRRHEVEQPEYDCGRDPP---DRDFP---TYPGEARP---EPRPYDSSRARQASGPH 478  
 Db 463 GRR-----GRKRPYEDRSIESDLYPGEFRNSFSNPQASSMMYETTGGRH 511

Qy 479 -----TITALYGAITSQLOLAHMART-----  
 Db 512 LSQTGPVTLNSLGAITSQLOKEVERLNGGNLPLISNAQSYYGPNMGHAPVYYSPPGT-- 569

Qy 516 HPHADTEPQAQPYPKAVYLPPPHIAPGPPGLSAVPPSYPPVATPGPAPPLHQPS 575  
 Db 570 HPTVSW-----PMGYVERMPMSTEGKTSTSNS---TVIPTPV-----601

Qy 576 PAHAHPPPPPPGPTPPPAASLQQPEAAGAL-----VMASSAAHVNVDTARAADLF 629  
 Db 602 -----SDPEAGRNPVPTATISQERSDGIQKESIEQSRTDNNAASAVGHTSDAGVDVF 655

Qy 630 VSOMM 634  
 Db 656 INQMM 660

RESULT 13  
 US-09-167-434-9  
 Patent No. 6008033  
 GENERAL INFORMATION:  
 APPLICANT: Abdel-Meguid, Sherin  
 APPLICANT: Qiu, Xiayang  
 APPLICANT: Culp, Jeffrey  
 APPLICANT: Debouck, Christine  
 APPLICANT: Janson, Chery A.  
 APPLICANT: Hoog, Susan S.  
 APPLICANT: Smith, Ward W.  
 TITLE OF INVENTION: No. 60080331 Proteases, Compositions Capable of

!TITLE OF INVENTION: Binding to said site, and Methods of Use Thereof  
 !NUMBER OF SEQUENCES: 18  
 !CORRESPONDENCE ADDRESS:  
 !ADDRESSEES: SmithKline Beecham Corporation  
 !STREET: 709 Swedeland Road - P.O. Box 1539  
 !CITY: King of Prussia  
 !STATE: Pennsylvania  
 !COUNTRY: USA  
 !ZIP: 19406-2799  
 !COMPUTER READABLE FORM:  
 !MEDIUM TYPE: Floppy disk  
 !COMPUTER: IBM PC compatible  
 !OPERATING SYSTEM: PC-POS/MS-DOS  
 !SOFTWARE: Patent in Release #1.0, Version #1.1.30  
 !CURRENT APPLICATION DATA:  
 !APPLICATION NUMBER: US/09/167,434  
 !FILING DATE:  
 !CLASSIFICATION:  
 !PRIORITY APPLICATION DATA:  
 !APPLICATION NUMBER: 08/853,755  
 !FILING DATE:  
 !PRIORITY APPLICATION DATA:  
 !APPLICATION NUMBER: US 60/022,470  
 !FILING DATE: 26-JUL-1996  
 !PRIORITY APPLICATION DATA:  
 !APPLICATION NUMBER: US 60/024,416  
 !FILING DATE: 21-AUG-1996  
 !PRIORITY APPLICATION DATA:  
 !APPLICATION NUMBER: US 60/030,901  
 !FILING DATE: 14-NOV-1996  
 !PRIORITY APPLICATION DATA:  
 !APPLICATION NUMBER: US 60/035,973  
 !FILING DATE: 21-JAN-1997  
 !PRIORITY APPLICATION DATA:  
 !APPLICATION NUMBER: US 60/039,191  
 !FILING DATE: 27-FEB-1997  
 !ATTORNEY/AGENT INFORMATION:  
 !NAME: Dinner, Dara L.  
 !REGISTRATION NUMBER: 33,680  
 !REFERENCE/DOCKET NUMBER: P50472-1  
 !TELECOMMUNICATION INFORMATION:  
 !TELEPHONE: 610-270-5022  
 !TELEFAX: 610-270-5090  
 !INFORMATION FOR SEQ ID NO: 9:  
 !SEQUENCE CHARACTERISTICS:  
 !LENGTH: 243 amino acids  
 !TYPE: amino acid  
 !STRANDEDNESS:  
 !TOPOLOGY: linear  
 !MOLECULE TYPE: protein  
 !US-09-167-434-9

Query Match 17.4%; Score 590.5; DB 3; Length 243;  
 Best Local Similarity 50.4%; Pred. No. 3.5e-18;  
 Matches 124; Conservative 34; Mismatches 79; Indels 8 11; Gaps 3;

Qy 1 MASAENPRLERAPLPPDRAVPIYVAGGLALYDGGDPPGELALDPDTVRAALPPENPLPINV 60  
 Db 1 MAAEADENCEA-----LYVAGIAYL-SKDEEELNTPEIVSALPPTSKINID 51

Qy 61 HRARCEYGRVIAVNNDRGPPFVGLIACVQLERVLETAASAAIFERGPALSREERLLYL 120  
 Db 52 HRKDCCVYGEVIAIIEIRGPPLFVYIPTCQPHAVLFEARHNNFGDRDSVSPLEALYL 111

Qy 121 ITNYLPSVSLSTKRCIDEVPDRTLPFHVVALCAIGRLGTVYDTSLDAIAFPFHLD 180  
 Db 112 VINYLPSVSLSRPSNEIDGNFPTVHALCVVGRGVTVNNDTPESSIEPPVLSM 171

Qy 181 ATREGYREAAEALALAGTWAQGEAAGAL-----VMASSAAHVNVDTARAADLF 240  
 Db 172 ESK--ARLLSLVVKDAGLNKWKVSSDKLAKVLLSTAVNNMLDRWDVAKRREAGIM 229

Qy 241 GHTYLOAS 248



```

; REFERENCE/DOCKET NUMBER: P50472-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5022
; TELEFAX: 610-270-5900
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 255 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-167-434-10

Query Match 17.4%; Score 590.5; DB 3; Length 255;
Best Local Similarity 50.0%; Pred. No. 3.7e-35;
Matches 124; Conservative 34; Mismatches 79; Indels 11; Gaps 3;
Qy 1 MASAEMRERLEAPLPDRAVPIYVAGFLALYDSDPGRLALDPTVRAALPPENPLPINVD 60
Db 1 MAAEADENCEA-----LYVAGLYALY-SKDEGELNITPEIVRSLAPPTSKIPINID 51
Qy 61 HRARCEVGRVLAVNNDPRGPPFFVGLIACVQLERVLETAASAAIFPERGPALSREERLLY 120
Db 52 HRKDCVYGEVIAITEDIRPFFGIVRCPLQHAYLFAAHNSNFFGNDTSPLERALYL 111
Qy 121 ITNYLPSPSLSTRRGDEVPPRTLFAHVALCAIGRLGTTVYDTSLDAAATAPFRHLDP 180
Db 112 VINYLPSTLSSCRSLSPNEIPQGNFFTHVALCVGRGVTVNYDCTPESSEIPEFRVLSM 171
Qy 181 ATREGVREAAEALALAGRTWAPGVFAITHPLSTAVNNNMRLDRWSLVAERRQAGIA 240
Db 172 ESK- -ARILSLVYDAGIWKWVSEDKLAKYLLSTAVNNMILRDRWDVAKRREAGIM 229
Qy 241 GHVYLOAS 248
Db 230 GHVYLOAS 237

```

Search completed: January 10, 2005, 19:20:04  
 Job time : 44 secs

| Result No. | Score | Query | Match | Length | DB ID                | Description          |
|------------|-------|-------|-------|--------|----------------------|----------------------|
| 1          | 3385  | 100.0 | 637   | 15     | US-10-623-429-9      | Sequence 9, Appli    |
| 2          | 2333  | 68.9  | 635   | 14     | US-10-214-932-86     | Sequence 86, Appli   |
| 3          | 351   | 10.4  | 536   | 17     | US-10-779-597-130995 | Sequence 30, Appli   |
| 4          | 294.5 | 8.7   | 534   | 16     | US-10-431-963-159995 | Sequence 159,995,    |
| 5          | 292   | 8.6   | 639   | 14     | US-10-368-493-3962   | Sequence 3962, Appli |
| 6          | 284   | 8.4   | 825   | 16     | US-10-408-765-2285   | Sequence 2285, Appli |
| 7          | 281.5 | 8.3   | 483   | 16     | US-10-431-963-191498 | Sequence 191498      |
| 8          | 279.5 | 8.3   | 569   | 16     | US-10-431-963-19216  | Sequence 149216,     |
| 9          | 275.5 | 8.1   | 426   | 16     | US-10-431-963-142229 | Sequence 142229,     |
| 10         | 273   | 8.1   | 506   | 16     | US-10-431-963-126337 | Sequence 126337,     |
| 11         | 271.5 | 8.0   | 521   | 16     | US-10-431-963-179715 | Sequence 179715,     |
| 12         | 270   | 8.0   | 606   | 16     | US-10-431-963-195106 | Sequence 195106,     |
| 13         | 268.5 | 7.9   | 373   | 16     | US-10-431-963-192104 | Sequence 149274,     |







RESULT 9  
 US-10-437-963-142229-9  
 US-10-437-963-142229 ; Sequence 142229, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
 ; Title of Invention: Rice Nucleic Acid Molecules and Other Molecules Associated with  
 ; File Reference: 38-21 (53221)B  
 ; Current Application Number: US/10/437,963  
 ; Current Filing Date: 2003-05-14  
 ; Number of SEQ ID NOS: 204966  
 ; SEQ ID NO: 142229  
 ; LENGTH: 426  
 ; TYPE: PRT  
 ; ORGANISM: Oryza sativa  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1) .. (569)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 ; FEATURE:  
 ; ORGANISM: Oryza sativa  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_4325C.1.pep  
 ; US-10-437-963-142276

Query Match 8.3%; Score 279.5; DB 16; Length 569;  
 Best Local Similarity 27.3%; Pred. No. 4.6e-09;  
 Matches 120; Conservative 23; Mismatches 150; Indels 147; Gaps 25;

Qy 257 ESAPAPERGYK------TGAGAMTSTPAAVYAROVAASSSSSSFPAP 306  
 Db 78 DERPPAPRQYRHPPPPQLRFGCPCCCCCCCCCCCCCCCCCCCCCCCC 132

Query Match 8.3%; Score 279.5; DB 16; Length 569;  
 Best Local Similarity 27.3%; Pred. No. 4.6e-09;  
 Matches 120; Conservative 23; Mismatches 150; Indels 147; Gaps 25;

Qy 307 ADNPVASSAA- PAPPPDGSYIWLIPASHYNOLVTGGSAPRHPPLTACGTLPAAGTY 364  
 Db 133 AHLPASAAGCPCPPXPPPA-----PSSTTTTATPPXPXPSAPPDPAPLPASVYC 185

Qy 365 GHPGAGP- SPHYPPPAAHPYPGML- -FAGPSPLEAQIAALYGAIAADRGAGSLPAAG 419  
 Db 186 PPPXPPAPPSSPPP- -XPXSSMTTTATPPPP-----RAVGSSRPPADPPAX- 233

Qy 420 DHGIRGSAKRRRHEQPEYDGRDPEP- -RDFPYYPGEARPEP- -PVDSR 467  
 Db 234 -----SPPPPXPSSPSAPSPAPSS- -PSSSPHPRPTQMPBPASSR 278

Qy 468 RAAROA-----SGPHETITALVGAVTSLOOBALMABR- THAPGYPVPGVPH 516  
 Db 279 RSTPRXSLSPWPWASPSPPH-----SSSSPSAASSSTARTRRPPPTPVKPTSS 328

Qy 517 -----PHA-----DTEPAQPPRYA-----KAYLPPPHI-----APPGPPH- 549  
 Db 329 TMVTPPARPSPSRPTSPPPP- -PAGLPLPLPLPPLPPLPAPPRPPLP 386

Qy 550 -SGAVPPPSYPPVAVTPGAPPLHQPS- -AH-AHPPP----- 583  
 Db 387 PPPSPPPSPPVNDLPPPPP- -PPPPRRRPRARRGGAEATPPRGGGVRDEDA 444  
 Qy 584 -PPSPTPPAASLQPEAP 602  
 Db 445 LPPPPPPPRPLPPLPPLP 464

RESULT 10  
 US-10-437-963-126337-7  
 US-10-437-963-126337 ; Sequence 126337, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua  
 APPLICANT: Cao, Yongwei  
 APPLICANT: Wu, Wei  
 APPLICANT: Boukharov, Andrey A.  
 APPLICANT: Barbazuk, Brad  
 APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With FILE REFERENCE: 38-21(53221)B  
 CURRENT APPLICATION NUMBER: US/10/437,963  
 CURRENT FILING DATE: 2003-05-14  
 SEQ ID NO: 204966  
 LENGTH: 506

TYPE: PRT  
 ORGANISM: *Oryza sativa*  
 FEATURE:  
 NAME/KEY: unsure  
 LOCATION: (1)..(521)  
 OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53221)B  
 CURRENT APPLICATION NUMBER: US/10/437,963  
 CURRENT FILING DATE: 2003-05-14  
 SEQ ID NO: 126337  
 LENGTH: 506

TYPE: PRT  
 ORGANISM: *Oryza sativa*  
 FEATURE:  
 NAME/KEY: unsure  
 LOCATION: (1)..(506)  
 OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_28894C.1.pep  
 US-10-437-963-126337

Query Match Score 8.1%; Score 273; DB 16; Length 506;  
 Best Local Similarity 29.2%; Pred. No. 1e-08;  
 Matches 109; Conservative 18; Mismatches 130; Indels 116; Gaps 19;

Qy 270 APGAMDTSPAAASVPAQVAVRA---RCVASSPAPADMN-----P 311  
 Db 30 SPIAVIASSVPAQCRVRAVSSPAALCVPLSSPSVPRXPSSPPPA 89

Qy 312 VSASGAPAPPDPGDSY-LWIPASHNQLVTCQSAPRH---PPLTACGLPAAGTVAYCHPGA 369  
 Db 90 XPRACAPPAPPDPD-----VV-----PPPSAXPAGPQFARRP-----P 125

Qy 372 SPHYP-----PPDA-----HPYPCGNLFLAGPSPLAQIAALVGATAADROAGGLPAAAG 419  
 Db 126 SPLRPLGGREGSEGPFAASRRPAPP-----ASPHPPPAR-----SGC 163

Qy 420 DHGIRGSACKRREHEQPEYDGRDDEPDRDFFYPPGEARPEP-RP-VDSRRAARQASGP- 476  
 Db 164 XEGEKKPPATPPTXPXPSDAVRSPDGDIRESPPSAGLPLRPLDEGGEPPLXP 223

Qy 477 -HETITALGAVTSILOQELAHMBARTHAPYGPYPPVYHHP-----HADDETPRQPR 529  
 Db 224 CARPPVRLPDAVR-----PSPDASPLPPPARSGGCGXEXPPXPPL 265

Qy 530 YPAKAVYLPPHIIAPPGPPLSAAVPPPSYPPVAVTPGAPPPLHQPSPAHAHPPPPGPT 589  
 Db 266 PPQ-----PPRLAQRBPP-----PPRECPFPFPFPFPFP-----PPFPFPPL 313

Qy 590 PPPAASLPLQPEAP 602  
 Db 314 PPPLAPRCEPPLP 326

RESULT 11  
 US-10-437-963-179715  
 Sequence 179715, Application US/10437963  
 Publication No. US20040123343A1  
 GENERAL INFORMATION:  
 APPLICANT: La Rosa, Thomas J.  
 APPLICANT: Kovalic, David K.  
 APPLICANT: Zhou, Yihua  
 APPLICANT: Cao, Yongwei  
 APPLICANT: Wu, Wei  
 APPLICANT: Boukharov, Andrey A.  
 APPLICANT: Barbazuk, Brad  
 APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With FILE REFERENCE: 38-21(53221)B  
 CURRENT APPLICATION NUMBER: US/10/437,963  
 CURRENT FILING DATE: 2003-05-14  
 NUMBER OF SEQ ID NOS: 204966  
 LENGTH: 606

TYPE: PRT  
 ORGANISM: *Oryza sativa*  
 FEATURE:  
 NAME/KEY: unsure  
 LOCATION: (1)..(606)  
 OTHER INFORMATION: unsure at all Xaa locations

OTHER INFORMATION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53221)B  
 TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With FILE REFERENCE: 38-21(53221)B



Search completed: January 10, 2005, 19:31:11  
 Job time : 150 secs

---

RESULT 15  
 US-10-425-115-231956  
 Sequence 231956, Application US/10425115  
 Publication No. US2004021427A1  
 GENERAL INFORMATION:  
 APPLICANT: La Rosa, Thomas J.  
 APPLICANT: Kovalic, David K.  
 APPLICANT: Zhou, Yihua  
 APPLICANT: Cao, Yongwei  
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants  
 FILE REFERENCE: 38-21(53222)B  
 CURRENT APPLICATION NUMBER: US/10/425,115  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 369326  
 SEQ ID NO 231956  
 LENGTH: 280  
 TYPE: PRT  
 ORGANISM: Zea mays  
 FEATURE:  
 NAME/KEY: unsure  
 LOCATION: (1)..(280)  
 OTHER INFORMATION: unsure at all xaa locations  
 FEATURE:  
 OTHER INFORMATION: Clone ID: MRT4577\_143137C.1.pep  
 US-10-425-115-231956

Query Match 7.9%; Score 267; DB 17; Length 280;  
 Best Local Similarity 27.8%; Pred. No. 1.3e-08;  
 Matches 84; Mismatches 84; Indels 128; Gaps 12;

Qy 304 PAPADNNPVSAGAPAPPGDGSYLMTPASHYINQVGTQSAAPRHPPLTACCLPAAGTVA 363  
 Db 65 PPPHQRTSPSPPPPSPG-----PPHPHPATPPXXPP-----100  
 Qy 364 YGHPGAGPSPHYBP-PPAHPYPGMLFAGPSPLEAQIAALVGIAAADRQAGGLPAAAGDHG 422  
 Db 101 --PPPPXXPHPPQPPNP-PPTTHPPPTP-----127  
 Qy 423 IRGSAKRRRHEVQPEIDCGRDEPDRDFPYGEARPRYHPHADTEPAPPRYPAKAVLPPPHI 482  
 Db 128 -----PPTTTHPPPTP-----GP-----152  
 Qy 483 LVGAVTSLQELAHMRARTHAPYGPYPPGVPHYHPP-----TTPPPPPPXPXPPPPPS 186  
 Db 153 -----PPPPXPNNPPPPH-----TPPPPPPXPXPPPPPS 186  
 Qy 543 APPGPPPLSGAVPPPSYPPVAVTPGPAPPBLHQPSPAHHP-PPPPGPTPPPAASLQPE 600  
 Db 187 SPDPPL---PSP--PMLPPSPHPPPTPPTP-----HPRQPPPPNPNTPPPT 239  
 Qy 601 AP 602  
 Db 240 PP 241

| Result No. | Score  | Query Match | Length | DB | ID       | Description        |
|------------|--------|-------------|--------|----|----------|--------------------|
| 1          | 3385   | 100.0       | 637    | 6  | AAE36428 | Aae36428 HSV-2 UL2 |
| 2          | 3385   | 100.0       | 637    | 8  | ADJ77800 | Adj77800 Herpes 61 |
| 3          | 3370.5 | 99.6        | 638    | 2  | AAR71017 | Aar71017 HSV-2 pro |
| 4          | 3359.5 | 99.2        | 642    | 2  | AAR72124 | Aar72124 HSV-2 str |
| 5          | 3351.5 | 99.0        | 636    | 2  | AAR64767 | Aar64767 HSV-2 pro |
| 6          | 3267.5 | 96.5        | 657    | 2  | AAW72192 | Aaw72192 HSV-2 str |
| 7          | 2333   | 68.9        | 635    | 2  | AAR71031 | Aar71031 HSV-1 UL2 |
| 8          | 2333   | 68.9        | 635    | 7  | ADL18166 | Adl18166 Human her |
| 9          | 2329   | 68.8        | 635    | 2  | AAR28634 | Aar28634 UL26 prot |
| 10         | 2326   | 68.7        | 635    | 2  | AAR28651 | Aar28651 UL26 prot |
| 11         | 2323.5 | 68.6        | 636    | 2  | AAR28636 | Aar28636 UL26 prot |
| 12         | 2322   | 68.6        | 635    | 2  | AAR28647 | Aar28647 UL26 prot |
| 13         | 2321   | 68.6        | 635    | 2  | AAR28652 | Aar28652 UL26 prot |
| 14         | 2319   | 68.5        | 626    | 2  | AAR28640 | Aar28640 UL26 prot |
| 15         | 2319   | 68.5        | 635    | 2  | AAR28650 | Aar28650 UL26 prot |
| 16         | 2318.5 | 68.5        | 626    | 2  | AAR28637 | Aar28637 UL26 prot |
| 17         | 2318   | 68.5        | 636    | 2  | AAR28635 | Aar28635 UL26 prot |
| 18         | 2318   | 68.5        | 635    | 2  | AAR28649 | Aar28649 UL26 prot |
| 19         | 2306   | 68.1        | 635    | 2  | AAR28648 | Aar28648 UL26 prot |
| 20         | 2241   | 66.2        | 615    | 2  | AAR28642 | Aar28642 UL26 prot |
| 21         | 2209   | 65.3        | 603    | 2  | AAR28641 | Aar28641 UL26 prot |
| 22         | 2017   | 59.6        | 514    | 2  | AAR28643 | Aar28643 UL26 prot |
| 23         | 1826   | 59.5        | 448    | 2  | AAW72048 | Aaw72048 HSV-2 str |
| 24         | 1826   | 53.9        | 331    | 2  | AAR64769 | Aar64769 HSV-2 ISP |
| 25         | 1820   | 53.8        | 329    | 6  | Aae36429 | Aae36429 HSV-2 UL2 |





| Matches | 634: Conservative  | 1: Mismatches | 2: Indels | 1: Gaps | 1: |
|---------|--|---------------|-----------|---------|----|
| Qy      | 1 MASAENRERLEAPLDRAPVIVAGFLALYSGDPGEIALLDDPTVRAALPENPLPINVD      | 60            |           |         |    |
| Db      | 5 MASAENRERLEAPLDRAPVIVAGFLALYSGDPGEIALLDDPTVRAALPENPLPINVD      | 64            |           |         |    |
| Qy      | 61 HRARCEVGRVLAVVNDPRGPFVGLIACVQLERVLETAASAAIFERRGPAISREERILYL   | 120           |           |         |    |
| Db      | 65 HRARCEVGRVLAVVNDPRGPFVGLIACVQLERVLETAASAAIFERRGPAISREERILYL   | 124           |           |         |    |
| Qy      | 121 ITNYLPSVSLSTKRGDDEVPLDRAFTAHVALCAIGRLGTVITYDTSLDAIAFPHRLDP   | 180           |           |         |    |
| Db      | 125 ITNYLPSVSLSTKRGDDEVPLDRAFTAHVALCAIGRLGTVITYDTSLDAIAFPHRLDP   | 184           |           |         |    |
| Qy      | 181 ATREGVREAAEELALAGTWAQGEALTHILLSTAVVNMMLRDMSLVAERRQGIA        | 240           |           |         |    |
| Db      | 185 ATREGVREAAEELAAGTWAQGEALTHILLSTAVVNMMLRDMSLVAERRQGIA         | 244           |           |         |    |
| Qy      | 241 GHTYLQSEKPKXWGAESAPAPERGKTYGKAPGMDTSPAAVASVAPAVARQVA-SSSS    | 299           |           |         |    |
| Db      | 245 GHTYLQSEKPKXWGAESAPAPERGKTYGKAPGMDTSPAAVASVAPAVARQVA-SSSS    | 304           |           |         |    |
| Qy      | 300 SSSFPAPADMNPPVSAAGAAPPGGDSYSLNIPASINYNQLTGQSAAPHPLTACGLPAA   | 359           |           |         |    |
| Db      | 305 SSSFPAPADMNPPVSAAGAAPPGGDSYSLNIPASINYNQLTGQSAAPHPLTACGLPAA   | 364           |           |         |    |
| Qy      | 360 GTVAYGHGPAGSPSPHIVPPPAHPYPMGLFAGSPSLEQIAALVGAIAADRQNGGLPAAAG | 419           |           |         |    |
| Db      | 365 GTVAYGHGPAGSPSPHIVPPPAHPYPMGLFAGSPSLEQIAALVGAIAADRQNGGLPAAAG | 424           |           |         |    |
| Qy      | 420 DHGIRGSAKRREHEVEQEYDGCGRDEPDRDPYPPGARPEPRPVDSRRAQASGPHE      | 479           |           |         |    |
| Db      | 425 DHGIRGSAKRREHEVEQEYDGCGRDEPDRDPYPPGARPEPRPVDSRRAQASGPHE      | 484           |           |         |    |
| Qy      | 480 ITALGVAVTSQELAHMRATHAPICPYPPNGPYTHPHADHETPAPPPRYPAKAYLPP     | 539           |           |         |    |
| Db      | 485 ITALGVAVTSQELAHMRATHAPICPYPPNGPYTHPHADHETPAPPPRYPAEAVYLPP    | 544           |           |         |    |
| Qy      | 540 PHIAPPGPPLSAGAVPPPSYPPVAVTPGPAPLHQPSDAHHPPPPQGPTPPAASLQP     | 599           |           |         |    |
| Db      | 545 PHIAPPGPPLSAGAVPPPSYPPVAVTPGPAPLHQPSDAHHPPPPQGPTPPAASLQP     | 604           |           |         |    |
| Qy      | 600 EAPGAEGALVNASSAAHVNVDTARAADLFLVSQMMGSR                       | 637           |           |         |    |
| Db      | 605 EAPGAEGALVNASSAAHVNVDTARAADLFLVSQMMGSR                       | 642           |           |         |    |
| Qy      | RESULT 5   |               |           |         |    |
| Qy      | AAR64767   |               |           |         |    |
| Qy      | ID AAR64767 standard; protein: 636 AA.                           |               |           |         |    |
| AC      | AAR64767;  |               |           |         |    |
| XX      | XX   |               |           |         |    |
| XX      | 25-MAR-2003 (revised)  |               |           |         |    |
| XX      | 19-JUL-1995 (first entry)  |               |           |         |    |
| DE      | HSV-2 protease, ICP35.   |               |           |         |    |
| KW      | Protease; HSV-2; protease-inhibitor; virucide; ICP35.            |               |           |         |    |
| XX      | Herpes simplex virus type 2.                                     |               |           |         |    |
| PD      | 22-DEC-1994.   |               |           |         |    |
| XX      | 25-MAY-1994;   |               |           |         |    |
| PA      | 94WO-US005920.   |               |           |         |    |
| XX      | XX   |               |           |         |    |
| XX      | 08-JUN-1993;   |               |           |         |    |
| PR      | 23-MAY-1994;   |               |           |         |    |
| PR      | 94US-00073819.   |               |           |         |    |
| XX      | XX   |               |           |         |    |
| XX      | W0429456-A2.   |               |           |         |    |
| XX      | W0429456-A2.   |               |           |         |    |
| PD      | XX   |               |           |         |    |
| XX      | 601 APGAEAGALVNASSAAHVNVDTARAADLFLVSQMMGSR                       |               |           |         |    |
| DB      | 600 APGAEAGALVNASSAAHVNVDTARAADLFLVSQMMGSR                       |               |           |         |    |
| XX      | 600  |               |           |         |    |
| XX      | AAW72192 standard; protein: 657 AA.                              |               |           |         |    |
| XX      | ID AAW72192  |               |           |         |    |
| XX      | RESULT 6   |               |           |         |    |

|                           |   |                                    |  |
|---------------------------|---|------------------------------------|--|
| AC                        | AAW72192;   | QY                                 | 360 GTVAYGHPGAGPSPHYPPPAHPYPMGLPAGPSPLEAQIAALVGAIAADRQAGGLPAAAG 419    |
| XX                        | 13-JAN-1999 (first entry)   | Db                                 | 361 GTVAYGHPGAGPSPHYPPPAHPYPMGLPAGPSPLEAQIAALVGAIAADRQAGGLPAAAG 420    |
| DE                        | HSV-2 strain SBS Contig ID 15 ORF#27 protein.                             | QY                                 | 420 DHGIRGSAKRRRHEVQPYDCRDPEPDRDFPYYPGEARPEPRPDVSRRAARQSGPHET 479      |
| XX                        |   | Db                                 | 421 DHGIRGSAKRRRHEVQPYDCRDPEPDRDFPYYPGEARPEPRPDVSRRAARQSGPHET 480      |
| KW                        | HSV-2 strain SBS; immunological response induction; therapy;              | QY                                 | 480 ITALVGAVTISLQOFLAHMRARTHAPYGPYPPVGPYHPPHADTEPAQPQPRYPKAVYLPP 539   |
| KW                        | antiviral identification; viral protein inhibitor.                        | Db                                 | 481 ITALVGAVTISLQOFLAHMRARTHAPYGPYPPVGPYHPPHADTEPAQPQPRYPKAVYLPP 540   |
| XX                        | Herpes simplex virus 2.   | QY                                 | 540 PHTAPPGPGLLSGAVPPSYPPVAVTPGPAPPLHOPSPAHAPPPPPGPPTPPDAAASLPP 599    |
| OS                        |   | Db                                 | 541 PHTAPPGPGLLSGAVPPSYPPVAVTPGPAPPLHOPSPAHAPPPPPGPPTPPDAAASLPP 600    |
| PN                        | WO9820016-A1.   | PF                                 | 31-OCT-1997; 97WO-US020016.  |
| XX                        |   | PR                                 | 04-NOV-1996; 96US-0030279P.  |
| PD                        | 14-MAY-1998.  | PR                                 | 09-JUN-1997; 97US-0049018P.  |
| XX                        |   | PA                                 | (SMIK ) SMITHKLINE BEECHAM CORP.                                       |
| PA                        |   | XX                                 | RESULT 7   |
| PI                        | Esser KM, Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB;         | ID                                 | AAR71031 standard; protein; 635 AA.                                    |
| PI                        | Leary JU;   | XX                                 | AAR71031;  |
| XX                        |   | AC                                 | AAR71031;  |
| DR                        | WPI; 1998-286847/25.  | XX                                 | XX   |
| N-PSDB;                   | AAV623176.  | DR                                 | 25-MAR-2003 (revised)  |
| XX                        | Herpes simplex virus type-2 sequences - useful in, e.g. prevention and    | DT                                 | 02-OCT-1995 (first entry)  |
| PT                        | treatment of infection or inducing immunological response in mammal.      | XX                                 | HSV-1 UL26 gene product.   |
| PT                        |   | DE                                 | HSV-1 UL26 gene product.   |
| XX                        |   | XX                                 | Herpes simplex virus; HSV-1 protease; capsid.                          |
| PS                        | Claim 10; Page 120; 740pp; English.                                       | XX                                 | Herpes simplex virus type 1.   |
| XX                        |   | OS                                 | Herpes simplex virus type 1.   |
| DR                        | W09506055-A1.   | XX                                 | W09506055-A1.  |
| XX                        |   | PN                                 | W09506055-A1.  |
| DR                        | 02-MAR-1995.  | XX                                 | XX   |
| XX                        |   | PD                                 | 02-MAR-1995.   |
| XX                        |   | XX                                 | XX   |
| PF                        | 19-AUG-1994;  | PF                                 | 94WO-US009303.   |
| XX                        |   | PR                                 | 20-AUG-1993; 93US-00110522.  |
| PR                        |   | PR                                 | 23-JUN-1994; 94US-00264537.  |
| XX                        |   | XX                                 | XX   |
| SQ                        | Sequence 657 AA;  | PS                                 | (SMIK ) SMITHKLINE BEECHAM CORP.                                       |
| CC                        | This sequence represents a Herpes simplex virus type-2 (HSV-2) protein    | PA                                 | PA   |
| CC                        | sequence of the invention. This sequence was isolated from a HSV-2 strain | XX                                 | XX   |
| CC                        | SBS (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.     | PI                                 | Dilella AG, Debouck CM;  |
| CC                        | Based on homology, this sequence is a capsid protein P40. The proteins    | XX                                 | XX   |
| CC                        | can be used for the treatment or prevention of disease, to induce an      | DR                                 | WPI; 1995-106803/14.   |
| CC                        | immunological response in a mammal or to identify inhibitors, activators  | DR                                 | N-PSDB; AAQ04678.  |
| CC                        | or novel antivirals. Antagonists of the proteins can be used to inhibit a | XX                                 | XX   |
| CC                        | viral polypeptide. The DNA sequence or a vector containing it can also be | PT                                 | New herpes simplex virus (HSV)-2 protease and capsid protein - used to |
| CC                        | used to induce an immunological response in a mammal                      | PT                                 | develop prods. for use in the diagnosis and treatment of HSV-2         |
| XX                        |   | PT                                 | infections. See also R71017-30. (Updated on 25-MAR-2003 to correct PN  |
| XX                        |   | XX                                 | field.)  |
| SQ                        | Sequence 657 AA;  | PS                                 | Disclosure; Page 35; 51pp; English.                                    |
| Query Match               | Score 32/7.5; DB 2; Length 657;   | PA                                 | (SMIK ) SMITHKLINE BEECHAM CORP.                                       |
| Best Local Similarity     | 96.5%; Pred. No. 1-4e-201;  | XX                                 | PA   |
| Matches 615; Conservative | 1; Mismatches 2; Indels 1; Gaps 1;  | PI                                 | PI   |
| QY                        | 1 MASAEMRERLEAPLDPDRAPVIAVNDPDPVRAALPENPLPIND 60                          | Dilella AG, Debouck CM;            | XX   |
| Db                        | 1 MASAEMRERLEAPLDPDRAPVIAVNDPDPVRAALPENPLPIND 60                          | DR                                 | WPI; 1995-106803/14.   |
| QY                        | 61 HRARCEVGVIAVNDPDPVFLVGLACVQLERVLETASAAIFERRGALSSEERLLY 120             | DR                                 | N-PSDB; AAQ04678.  |
| Db                        | 61 HRARCEVGVIAVNDPDPVFLVGLACVQLERVLETASAAIFERRGALSSEERLLY 120             | XX                                 | XX   |
| QY                        | 121 ITNYLPSVSLSTKRGDEPPDRTLFARVALCAGRRLGTTYDTSIDAAIFAPRFLDP 180           | PT                                 | PT   |
| Db                        | 121 ITNYLPSVSLSTKRGDEPPDRTLFARVALCAGRRLGTTYDTSIDAAIFAPRFLDP 180           | PT                                 | PT   |
| QY                        | 181 ATREGVRRRAAEALAGRTWAGVEAHTLTLSTAVNMMLDRSLSVAAERRQGIA 240              | CC                                 | CC   |
| Db                        | 181 ATREGVRRRAAEALAGRTWAGVEAHTLTLSTAVNMMLDRSLSVAAERRQGIA 240              | CC                                 | CC   |
| QY                        | 241 GHTYLQASRKFKIGAEASAPAPERYKTCGAGMTSPASVYDPAQIVAVRQYVA-SSS 299          | CC                                 | CC   |
| Db                        | 241 GHTYLQASRKFKIGAEASAPAPERYKTCGAGMTSPASVYDPAQIVAVRQYVA-SSS 300          | CC                                 | CC   |
| QY                        | 300 SSSFPAPADMNPVPSASGAAPPGFDGSYLWIAASHYMNOLYTCOSAAPHPLTACGILPA 359       | Query Match                        | 1 MASAEMRERLEAPLDPDRAPVIAVNDPDPVRAALPENPLPIND 60                       |
| Db                        | 301 SSSFPAPADMNPVPSASGAAPPGFDGSYLWIAFPHTNQLVTQSAPHPLTACGILPA 360          | Best Local Similarity              | 68.9%; Score 23/33; DB 2; Length 635;                                  |
| QY                        |   | Mismatches 30; Indels 120; Gaps 7; | Pre. No. 1-4e-141;   |
| Db                        |   | Matches 477; Conservative 30;      | Matches 477; Conservative 30;  |

|          |                      |   |     |             |               |  |
|----------|----------------------|---|-----|-------------|---------------|--|
| Qy       | 61                   | HRARCEYGRVLAVNDRGGPFFYGLIACVQLERVLETASAATFERRGSPALSREERLLYL   | 120 | XX          | PS            | Disclosure; SEQ ID NO 86: 214pp; English.  |
| Db       | 61                   | HRAGCEYGRVLAVNDRGGPFFYGLIACVQLERVLETASAATFERRGSPALSREERLLYL   | 120 | XX          | CC            | The present invention describes a chimeric protein comprising at least one signal probe that has a trafficking signal targeting to a subcellular organelle and at least one proteolytic cleavage site for a protease. The chimeric protein is constructed, so that: (a) the trafficking signals of all the signal proteins are inactivated by linking the proteolytic site or a signal masking protein through the proteolytic site to the N- or C- terminus of the signal proteins, and so the chimeric protein is present in cytosol; (b) the trafficking signal of at least one signal protein is activated when the proteolytic cleavage site is cleaved by the protease, and as a result at least one fragment protein that includes the activated signal protein is transported to a subcellular organelle; and (c) the chimeric protein is labelled with at least one fluorescent protein and the position and intensity distribution of the fluorescent label signal in the cell is altered depending on the cleavage by the protease. Also described: (1) a recombinant gene comprising the nucleic acid sequence encoding the chimeric protein which is constructed to express the chimeric protein in a cell; (2) a cell transformed with the recombinant gene or vector; (3) analysing the activity of a protease detecting a protease inside a cell; (4) screening protease inhibitors in vivo; (5) a system for sequencing a nucleic acid comprising the sequence encoding the chimeric protein for detecting protease activity in a cell; (6) a vector comprising the nucleic acid; (8) a kit for detecting a protease inside a cell comprising the chimeric protein or the vector; (9) detecting a protease inside cell or tissue; and (10) detecting a protease inhibitor in vivo. The chimeric protein is useful for detecting protease inhibitors inside the cell or tissue. The present sequence represents HSV-1 protease, which is used in the exemplification of the present invention. |
| Qy       | 121                  | ITNYLPVSLSTKRRGDEVPDTRLFAHVALCAIGRRGTTVYDTSLDAATAPFRHLDP      | 180 | CC          | CC            | CC   |
| Db       | 121                  | ITNYLPVSLSTKRRGDEVPDTRLFAHVALCAIGRRGTTVYDTGLDATAATAPFRHLDP    | 180 | CC          | CC            | CC   |
| Qy       | 181                  | ATREGVIREAAEELALAGRITWAPGEVTLHTLSTAVNNMMLDRWSLVAERROAGIA      | 240 | CC          | CC            | CC   |
| Db       | 181                  | ASREGARRLAEEELALSGRTWAPGEVTLHTLSTAVNNMMLDRWSLVAERROAGIA       | 240 | CC          | CC            | CC   |
| Qy       | 241                  | GHTYLOQASEKFIKWGAESAPAPERYGKTKGPGAMDTSPAAVSVAPO----VAVTRQVAS  | 296 | CC          | CC            | CC   |
| Db       | 241                  | GHTYLOQASEKFIKWGAEPVSAPARGYKNGAPESTDIPGTSIAAPQGDRCPITRQGVYL   | 300 | CC          | CC            | CC   |
| Qy       | 297                  | SSSSSFPAADMNPVSAGAPAPPGPGDSYLIWIPASHYNVOLVTGOSAPRHPPLTACGL    | 356 | CC          | CC            | CC   |
| Db       | 301                  | S-----PVLPMPNPVPVPSGTAPAPEDGSYLIWIPASHYNVOLVTGOSAPRHPPLTACGL  | 354 | CC          | CC            | CC   |
| Qy       | 357                  | P-AAGTVAHYGHCAAGPSHYPHYPGMLFAGPSPLEAQIAALVCAIAARQAGGLP        | 415 | CC          | CC            | CC   |
| Db       | 355                  | PAAGAGSVAHYGHAGLSQLHYPHYPVAHQPGVLFSGSPSLEAQIAALVCAIAARQAGQP   | 414 | CC          | CC            | CC   |
| Qy       | 416                  | AAAGDHGIGRGSAGRGRHVEQPDYDGFYTPGKAESEPEPVDPSREAAARQASG         | 475 | CC          | CC            | CC   |
| Db       | 415                  | -AAGDPGTVRGSGKGRRYEAQPSSESYCQDOPDADYPPGKGEARQAPRGVDSRAARHSPG  | 473 | CC          | CC            | CC   |
| Qy       | 476                  | PHETITIALVGATVTSLOOFLAHMBARTHAPYGPYPPGPYHHPADTE-TPAQOPPRYPAKA | 534 | CC          | CC            | CC   |
| Db       | 474                  | TNBTITIALMGATVTSLOOFLAHMBARTHAPYGPYPPGPYHHPADTE-TPAQOPPRYPAKA | 533 | CC          | CC            | CC   |
| Qy       | 535                  | VYLPPIPHIATPGPPLSGA--VPPPSYPPVAVTPGPAPPLHQPSPAHAHPPPPPGPTDPP  | 592 | XX          | XX            | XX   |
| Db       | 534                  | VYRPPPHSAPYGPQGPASHAPYPPVAPAACCPGPPPP--POPSTQTRAPLPTEPAPPP    | 590 | XX          | XX            | XX   |
| Qy       | 593                  | AASLPQPQPEAPGAAGALVNASSAAHVNYDTARAADLFLVFSQMMGSSR             | 637 | XX          | XX            | XX   |
| Db       | 591                  | AKTGSQPEASNAEAGLVNASSAAHVNYDTARAADLFLVFSQMMGAR                | 635 | XX          | XX            | XX   |
| RESULT 8 |                      |   |     |             |               |  |
| Qy       | ADL18166             | Human herpesvirus 1 protease protein SEQ ID NO:86.            | DT  | 06-MAY-2004 | (first entry) |  |
| Db       | ID ADL18166          | standard; protein: 635 AA.                                    | XX  | XX          | XX            |  |
| Qy       | AC ADL18166;         | Human herpesvirus 1.  | XX  | XX          | XX            |  |
| Db       | XX                   | WO2003014381-A1.  | XX  | XX          | XX            |  |
| Qy       | PD 20-FEB-2003.      |   | XX  | XX          | XX            |  |
| Db       | 08-AUG-2002;         | 2002WO-KR001515.  | XX  | XX          | XX            |  |
| Qy       | 10-AUG-2001;         | 2001KR-00048123.  | XX  | XX          | XX            |  |
| Db       | (AHRA-)              | AHRAM BIOSYSTEMS INC.   | XX  | XX          | XX            |  |
| Qy       | WPI, 2003-255596/25. |   | XX  | XX          | XX            |  |
| Db       | N-PSDB;              | ADL18165.   | XX  | XX          | XX            |  |
| Qy       | 476                  | PHEITALVGATVTSLOOFLAHMBARTHAPYGPYPPGPYHHPADTE-TPAQOPPRYPAKA   | 534 | PT          | PT            | New chimeric protein, useful for detecting protease inhibitors inside the cell or tissue.  |
| Db       | 474                  | TNETITALMGATVTSLOOFLAHMBARTHAPYGPYPPGPYHHPADTE-TPAQOPPRYPAKA  | 533 | PT          | PT            | New chimeric protein, useful for detecting protease inhibitors inside the cell or tissue.  |

|          |  |    |  |
|----------|--|----|--|
| Qy       | 535 VILPPPHIAPPGPPLISGA - VPPPSYPPVAVTPGPAPPQHOPSPAHAHPPPPGPTPP 592  | Db | 61 HRACEVGRVLAVIDPQGPFFVGLIACVQLBRVLETAASAAIFERRGPEPLSREERLLYL 120   |
| Db       | 534 VYRPPPHSAPYGPQGPASHAPTPYAAPAACPPGPPP - - PCPSQTQAPLPTEPAPP 590   | Qy | 121 ITNYLPSVSLSTKRGDVPVPPDRTLFAHVALCAGRRGGTIVYDTSDDAAIAAPPRHLDP 180  |
| Qy       | 593 AASLFQPEAFGAEGALVNASSAAHVNVTARADLFYSQMMGSR 637   | Db | 121 ITNYLPSVSLATKRGDRAHPDRTLFAHVALCAGRRGGTIVYDTSDDAAIAAPPRHLDP 180   |
| Db       | 591 AATGSQPEASNAEAGALVNASSAAHVDTARADLFYSQMMGAR 635   | Qy | 181 ATREGVREAAEALLAGRTWAPGEALTHTLISTAVNNMLDRWSIVAEERRQAGIA 240   |
| RESULT 9 |  |    |  |
| AAR28634 | standard: protein; 635 AA.   | Db | 181 ASNEGARRLAEEALSGRTWAPGEALTHTLISTAVNNMLDRWSIVAEERRQAGIA 240   |
| ID       | AAR28634 (revised)   | Qy | 241 GHTYLOASEKFKWIWGAESAPAPERGKTYKGAPGAMDTSPASVPAQ - - VAVRAQYAS 296   |
| XX       | DT 25-MAR-2003 (first entry)   | Db | 241 GHTYLOASEKFKWIWGAESAPAPERGKTYKGAPGAMDTSPASVPAQ - - VAVRAQYAS 296   |
| AC       | XX   | Qy | 297 SSSSSFAPADKNPVSSAGAAPPEDGSYWIAPSHYNQLVGQAPSTDIPPGSIAAAPPQDRCPIVQRGVAL 300  |
| DE       | UL26 protease.   | Db | 301 S - - - - - PVLPKPNPVPSGTAPAPPGDGSYWIAPSHYNQLVGQAPSTDIPPGSIAAAPPQDRCPIVQRGVAL 304  |
| XX       | UL26: open reading frame; ORF; protease; herpes simplex virus; HSV; capsid protein; ICP35; ICP35 c; ICP35 d; ICP35 e; ICP35 f; functional domain.  | Qy | 357 P-AAGTVAYGHPGAGSPHYPYPPPAHPYPPGMLPAGSPSLEQTAALVGAIAADRGAGGLP 415   |
| KW       | KW   | Db | 355 PAAGGSVAGYGPQGPAGLSQHYPHPVAHQTPGVLPSGSPLEAQTAALVGAIAADRGAGGP 414   |
| KW       | KW   | Qy | 416 AAQGDHGRGSAKRRHVEQPEYDCGDEDPDFPYGEARPEPVDPSRDAARQASG 475   |
| XX       | OS Herpes simplex virus type 1.  | Db | 415 -ARGDPGVGTRGSKRRYEAQPSSESYCDOEPDADPYPPGEARQAPRGVDSRAARHSPG 473   |
| XX       | EP514830-A2.   | Qy | 476 PHETITIALVGATVSLQQQLAHMNRARTHAYGPYPPVGYHHPDADE - TPAQGPYRYPAKA 534   |
| XX       | 25-NOV-1992.   | Db | 474 TNETTTALMGATVSLQQQLAHMNRARTSAYGMVTPVHTRPQVGEPEPTTTHPALCPPPA 533  |
| PD       | XX   | Qy | 535 VYLPPIPHIAPPGPPLS - - GAVPPPSYPPVAVTPGPAPPQHOPSPAHAHPPPPGPTPP 592  |
| PF       | 19-MAY-1992;   | Db | 534 VYRPPPHSAPYGPQGPASHAPTPYAAPAACPPGPP - - PCPSQTQAPLPTEPAPP 590  |
| XX       | 92EP-00108420.   | Qy | 593 AASLPQPEAPGABAGALVNASSAAHVNDTARADLFYSQMGSR 637   |
| PR       | 24-MAY-1991;   | Db | 591 AATGSQPEASNAEAGALVNASSAAHVNDTARADLFYSQMGAR 635   |
| PR       | 07-FEB-1992;   | XX | RESULT 10  |
| XX       | 92IS - 00832855.   | ID | AAR28651 standard; protein; 635 AA.  |
| (ARCH-)  | ARCH DEV CORP.   | XX | XX   |
| PI       | Roizman B, Liu F;  | AC | AAR28651;  |
| XX       | DR 1992-391444/48.   | XX | XX   |
| DR       | N-PSDB; AAQ30736.  | DT | 25-MAR-2003 (revised)  |
| XX       | New herpes serine proteases and correxp. nucleic acid sequences - for detection, prevention and treatment of infection caused by HSV, Epstein-Barr, Varicella-Zoster and CMV cytomegalo-virus.   | DT | 24-MAR-1993 (first entry)  |
| PT       | Disclosure: Fig 1B; 66pp; English.   | XX | XX   |
| PS       | This sequence is encoded by the UL26 open reading frame (ORF) and it is a protease. The UL26 ORF is contained within a section of the herpes simplex virus (HSV) type 1 genome for the same HSV capsid proteins designated ICP35. The UL26 transcription initiation site is used as the start point for measurements of distance in this section of the genome. The UL26 protease and ICP35 precursor proteins, ICP25 c and d, may be cleaved by the UL26 protease. Cleavage of ICP35 c and d gives ICP15 and f respectively. Cleavage specificity of the UL26 protease may be altered by mutations within the sequence. These mutations may be insertions, deletions or substitutions which affect the functional domains of the proteases. See also AAR28635-52. (Updated on 25-MAR-2003 to correct PN | DS | UL26 protease substitution mutant KK.  |
| CC       | Sequence 635 AA;   | XX | XX   |
| CC       | SQ   | XX | UL26: open reading frame; ORF; protease; herpes simplex virus; HSV; capsid protein; ICP35; ICP35 c; ICP35 d; ICP35 e; ICP35 f; functional domain; insertion; deletion; substitution. |
| CC       | Query Match Best Local Similarity 68.8%; Score 2329; DB 2; Length 635;   | XX | XX   |
| CC       | Matches 476; Conservative 30; Mismatches 121; Indels 18; Gaps 7;   | XX | XX   |
| Qy       | 1 MASAEMREFLEAPLIDPDRAPVPIVAGFLALYDSDGPGLELDPPDTVRAALPPNPLPNVD 60  | PR | 19-MAY-1991; 92EP-00108420.  |
| Db       | 1 MAADAPGDRMEEPIPDRAPVPIVAGFLALYDSDGPGLELDPPDTVRAALPPNPLPNVD 60  | PR | 24-MAY-1991; 91US-00705814.  |
| Qy       | 61 HRACEVGRVLAVIDPQGPFFVGLIACVQLERVLETAASAAIFERRGPEPLSREERLLYL 120   | PR | 07-FEB-1992; 92US-00832855.  |
| PI       | Roizman B, Liu F;  | PA | (ARCH-) ARCH DEV CORP.   |
| PI       |  | XX |  |

XX WPI: 1992-391444/48.  
XX  
XX New herpes serine proteases and corresp. nucleic acid sequences - for  
XX detection, prevention and treatment of infection caused by HSV, Epstein-  
XX Barr, Varicella-Zoster and CMV cytomegalo-virus.  
XX  
XX 38 Disclosure; Table 1; 66pp; English.  
XX  
XX The sequences given in AAB28635-52 are mutants of the protease encoded by  
XX the UL26 open reading frame (ORF). The UL26 ORF is contained within a  
XX section of the herpes simplex virus (HSV) type 1 genome for the family of  
XX HSV capsid proteins designated ICP35. The UL26 transcription initiation  
XX site is used as the start point for measurements of distance in this  
XX section of the genome. The UL26 protease and ICP35 precursor proteins,  
XX ICP35 c and d, may be cleaved by the UL26 protease. Cleavage of ICP35 c  
XX and d gives ICP35 e and f respectively. Cleavage specificity of the UL26  
XX protease is altered by the inclusion of these mutations. These mutations  
XX include insertions, deletions or substitutions which affect the functional  
XX domains of the protease. (Updated on 25-MAR-2003 to correct PN  
XX field.)

| Sequence 635 AA: |                       |  |                     |            |             |   |
|------------------|-----------------------|--|---------------------|------------|-------------|---|
| Query            | Match                 | Score  | DB 2;               | Length     | DB 1;       | 7 |
| Py               | Best Local Similarity | 68.7%  | Score 2326;         | DB 2;      | Length 635; |   |
| Py               | Matches 475;          | 73.6%  | Pred. No. 3..9e-11; |            |             |   |
| Py               | Conservative          | 31;  | Mismatches 121;     | Indels 18; | Gaps        |   |
| Py               | 1                     | MASAMEMBERLEAPLDPDAVPIVAGFLALYDSDCPGEIALDDDTVALPPNLPINV                    | 60                  |            |             |   |
| Py               | 1                     | MAADAPGDRMEEBLPDRAPIVVAGFLALYDSDGSEALDDPTVALPPNLPINV                       | 60                  |            |             |   |
| Py               | 61                    | HRARCEYGRVLAVVNDPRGPFFVGLIACVQLERVLETAASAAIFERRGPALSREERLLYL               | 120                 |            |             |   |
| Py               | 61                    | HRACCEYGRVLAVVNDPRGPFFVGLIACVQLERVLETAASAAIFERRGPPLREERLLYL                | 120                 |            |             |   |
| Py               | 121                   | ITNYLPVSYSLSKTRGRDEVPPDRTLFVAHVALCIGRLGTTIVYDTSLDAATAPFRHLDP               | 180                 |            |             |   |
| Py               | 121                   | ITNYLPVSYSLSATKRLGGEAHPDRTLFVAHVALCIGRLGTTIVYDTGIDAATAPFRHLDSP             | 180                 |            |             |   |
| Py               | 181                   | ATREGVREAAEALALAGRTVAPGVEALTHTLLSTAVNNMLRDMSLVAERRQGIA                     | 240                 |            |             |   |
| Py               | 181                   | ASREGARRLAAEALASGRTVAPGVEALTHTLLTAAVNNMLRDMSLVAERRQGIA                     | 240                 |            |             |   |
| Py               | 241                   | GHTYLOASEKEFKIWKGAESAPAPERGYKTGAPGAMDTSPAASVPAQO---VAVRAROYAS              | 296                 |            |             |   |
| Py               | 241                   | GHTYLOASEKEFKIWKGAESAPARGYKNGAPESTDIPGSTIAAQPQGDRCPIVRQGVAL                | 300                 |            |             |   |
| Py               | 297                   | SSSSSSPAPADMNPVSASGADAPPCCGSKLWIPASHYHNLVGTGOSAPRHPPLTACGL                 | 356                 |            |             |   |
| Py               | 301                   | S-----PVLPMPNPMPVTSGTAAPPCCGSKLWIPASHYHNLVGTGOSAPRHPPLTACGL                | 354                 |            |             |   |
| Py               | 357                   | P-AAGTYAGHPCAGPSPHYPPEPAHPYPMGLFAGPSPLEAQIAALYGAIAADRAQAGGLP               | 415                 |            |             |   |
| Py               | 355                   | PAAGAGTAYGPHCAGPSPHYPPEPAHPYPMGLFAQHYPVAHQPGVIFSGPSPLEAQIAALYGAIAADRAQAGQP | 414                 |            |             |   |
| Py               | 416                   | AAAGDHGTRGSAKRREVEQPEYDYGCRDEPDRDPYYPGEARPEPPVDSRAARASG                    | 475                 |            |             |   |
| Py               | 415                   | -AAGDPGEYRGSGKRRYEAQPSSEYCYQDPEADYPYYPGEARPGCVDSRAARHSPG                   | 473                 |            |             |   |
| Py               | 476                   | PHETITALGVATSLQQLAMMRATHAATPYGPVPPVSPYHHPDAH-TPAQDPYRYPKA                  | 534                 |            |             |   |
| Py               | 474                   | TNETITALMGAVTSLQQLAMMRATSAQYGMNTPYAHYRQVQEPEPTTHPALCPPEA                   | 533                 |            |             |   |
| Py               | 535                   | VYLPDPHITAAPPGFLS--GAVPPPSYPPEVATPGPAPPLHOPSPAHAHPPPPGPTPPP                | 592                 |            |             |   |
| Py               | 534                   | VYRPPPSAPYGPQGPDSHAPTPYPAACCPGPPP---BCPSTOTRPLPTEPAFPP                     | 590                 |            |             |   |
| Py               | 593                   | AASLQPQEAQGAQALYNASSAAHNVDTARAADLITYSOMGSR                                 | 637                 |            |             |   |
| Py               | 591                   | AATGSGOPAESNAEAGALYNASSAAHNVDTARAADLITYSOMGSR                              | 635                 |            |             |   |

RESULT 11

|                 |     |                      |                    |                  |                     |
|-----------------|-----|----------------------|--------------------|------------------|---------------------|
| AAR28636        | ID  | AAR28636             | standard           | protein          | 636 AA.             |
| XX              | AC  | AAR28636;            |                    |                  |                     |
| XX              | XX  |                      |                    |                  |                     |
| DT              | DT  | 25-MAR-2003          | (revised)          |                  |                     |
| DT              | DT  | 24-MAR-1993          | (first entry)      |                  |                     |
| XX              | XX  |                      |                    |                  |                     |
| DE              | DE  | UL26                 | protease           | insertion mutant | J.                  |
| XX              | XX  |                      |                    |                  |                     |
| KW              |     | UL26                 | open reading frame | ORF              | proto               |
| KW              |     |                      | capsid protein     | ICP5             | ICP                 |
| KW              |     |                      | functional domain  | 35 c;            | 35                  |
| XX              | XX  |                      | insertion          | insertion        | delet               |
| OS              | XX  |                      |                    |                  |                     |
| XX              | XX  |                      |                    |                  |                     |
| PH              |     | Herpes simplex virus | type 1.            |                  |                     |
| XX              | XX  |                      |                    |                  |                     |
| Key             |     |                      |                    |                  | Location/Qualifiers |
| Misc-difference | 515 |                      |                    |                  | /note= "Insertion m |
| FT              | FT  |                      |                    |                  |                     |
| FT              | FT  |                      |                    |                  |                     |
| XX              | XX  |                      |                    |                  |                     |

181 ATREGVRREAAELALAGRTWAPGVVEALTHLLSTAVNNMMRLDRNSLVAERRRQAGIA

|           |  |  |
|-----------|--|--|
| Db        | 181 ASREGARRLAAEAELASGRWTWPGVEALTHLTLISTAVNMMLDRWSLVAERRQAGIA 240  | CC The sequences given in AAR28635-52 are mutants of the protease encoded by the UL26 open reading frame (ORF). The UL26 ORF is contained within a section of the herpes simplex virus (HSV) type 1 genome for the family of HSV capsid proteins designated ICP35. The UL26 transcription initiation site is used as the start point for measurements of distance in this section of the genome. The UL26 protease and ICP35 precursor proteins, ICP35 c and d, may be cleaved by the UL26 protease. Cleavage specificity of the UL26 protease is altered by the ICP35 e and f respectively. Cleavage specificity of the protease, including insertions, deletions or substitutions which affect the functional domains of the protease. (Updated on 25-MAR-2003 to correct PN field.) |
| Qy        | 241 GHTYLOASKEFKIWAESAESAPERYKTKGAPGMDTSPASVPAFQ---VAVARQYVS 296   | CC   |
| Db        | 241 GHTYLOASKEFKIWAESAESAPERYKTKGAPGMDTSPASVPAFQ---VAVARQYVS 296   | CC   |
| Db        | 241 GHTYLOASKEFKIWAESAESAPERYKTKGAPGMDTSPASVPAFQ---VAVARQYVS 300   | CC   |
| Qy        | 297 SSSSSFPAPADMNPPVSAASGAAPAPPGGDSYLWTPASHYNOLVGTOSAPRHPPLTACGL 356   | CC   |
| Db        | 301 S-----PVLPMPMNVPPTSTPAPAPPGGDSYLWTPASHYNOLVGTOSAPRHPPLTACGL 354  | CC   |
| Qy        | 357 P-AAGTVYAGHPGAGPSPHYPPPPAHPPYCMGLFAGPSPLPEAQIAALVGAIAADRQAGGP 415  | CC   |
| Db        | 355 PAAAGSVAVGPHGAGLSQSYHPPHYAHQYCVLFSGSPSPEAQIAALVGAIAADRQAGGP 414  | CC   |
| Qy        | 416 AAAGDHGIRSSAKRERRHETEQPEYDGRDEPDRDPYVGEARBPBPVDSRRAARQAGS 475  | CC   |
| Db        | 415 -AAGDPGVRSGSKRRYEAQPSSESYCDQEDPDAQDPPYVGEARBPQGDSRRAARHSG 473  | CC   |
| Qy        | 476 PHEITIALVGAVTSILOQELAHMRARTHAYGPYPPVGPYTHPHADTEP---AQPPRYPAK 533   | CC   |
| Db        | 474 TNETITALMGAVTSLQOELAHMRARTSAPGMYTEVAHTRPQXQVGEPEPTTHALCPB 533  | CC   |
| Qy        | 534 AVYLPAPPHTAPPGPPLS---GAVPPPSYPPVATPGPAPPLHOPSPAHAHPPPPPGPTPP 591   | CC   |
| Db        | 534 AVYRPAPPHASPYGPQGPSSHAFTPPYPAACPCGPPPP---PCPSQTQRAPLPTEPAPP 590  | CC   |
| Qy        | 592 PAASLDPQPEAAGGAEGALVNSAAHVNDTARAADLFVTSQMMGSR 637  | CC   |
| Db        | 591 PAATGSQPEASNAEAGALVNSAAHVNDTARAADLFVTSQMMGAR 636   | CC   |
| RESULT 12 |  |  |
| ID        | AAR28647 standard; protein: 635 AA.  |  |
| AC        | AAR28647;  |  |
| XX        |  |  |
| DT        | 25-MAR-2003 (revised)  |  |
| DT        | 24-MAR-1993 (first entry)  |  |
| DE        | UL26 protease substitution mutant GG.  |  |
| XX        | UL26; open reading frame; ORF; protease; herpes simplex virus; HSV;  |  |
| KW        | capsid protein; ICP35 c; ICP35 d; ICP35 e; ICP35 f;  |  |
| KW        | functional domain; insertion; deletion; substitution.  |  |
| XX        | Herpes simplex virus type 1.   |  |
| OS        |  |  |
| XX        | Location/Qualifiers  |  |
| FT        | Misc-difference 7 .9   |  |
| FT        | /note= "Substitution CDR > SRT"  |  |
| XX        | EP514830-A2.   |  |
| XX        | 25-NOV-1992.   |  |
| PP        | 19-MAY-1992; 92EP-00108420.  |  |
| XX        |  |  |
| PR        | 24-MAY-1991; 91US-00705814.  |  |
| PR        | 07-FEB-1992; 92US-00832855.  |  |
| XX        |  |  |
| (ARCH-)   | ARCH DEV CORP.   |  |
| XX        |  |  |
| PI        | Roizman B, Liu F;  |  |
| XX        | WPI: 1992-391444/48.   |  |
| DR        |  |  |
| XX        | New herpes serine proteases and correxp. nucleic acid sequences - for detection, prevention and treatment of infection caused by HSV, Epstein-Barr, Varicella-Zoster and CMV cytomegalo-virus. |  |
| PT        |  |  |
| PT        |  |  |
| XX        | Disclosure; Table 1; 65pp; English.  |  |
| PS        |  |  |
| XX        |  |  |
| DE        | UL26 protease substitution mutant LL.  |  |
| RESULT 13 |  |  |
| ID        | AAR28652 standard; protein: 635 AA.  |  |
| XX        |  |  |
| AC        | AAR28652;  |  |
| XX        |  |  |
| DT        | 25-MAR-2003 (revised)  |  |
| DT        | 24-MAR-1993 (first entry)  |  |

|  |  |           |   |
|--|--|-----------|---|
| XX   | UL26; open reading frame; ORF; protease; herpes simplex virus; HSV;        | Qy        | 357 P-AAGTVAAGHPGAGGSPPHYPPPPAPYPGMLFAGGSPLAQIAALVGAIAADRGGLP 415   |
| KW   | capsid protein; ICP35; ICP35 c; ICP35 d; ICP35 e; ICP35 f;                 | Db        | 355 PAAAGGVAAGPHGAGISQHYPHVAQYPGMLFAGGSPLAQIAALVGAIAADRGGLP 414     |
| KW   | functional domain; insertion; deletion; substitution.                      |           |   |
| XX   | Herpes simplex virus type 1.   |           |   |
| XX   | Key  |           |   |
| FH   | Location/Qualifiers  |           |   |
| FT   | Misc-difference 34 /label= D34A  | Qy        | 416 AAAGDGGIRGSAKRRRHEVEPEYDGRDDEPDRDFFYPPGARPEPRPVDSRRAAHQASG 475  |
| FT   | EP514830-A2.   | Db        | 415 -AAGDGGIRGSAKRRRHEVEPEYDGRDDEPDRDFFYPPGARPEPRPVDSRRAAHQASG 475  |
| XX   | 25-NOV-1992.   | Qy        | 476 PHETITALVGAVTSLOQELAHMRTHAPYNGPYPPVGYHYPHADTE-TPAQPPRVPYAKA 534 |
| XX   | XX   | Db        | 474 TNETITALMGAVTSLOQELAHMRTHAPYNGPYPPVGYHYPHADTE-TPAQPPRVPYAKA 533 |
| XX   | 19-MAY-1992;   | Qy        | 535 VYLPPPHIAPPGPPLS -GAVPPPSTPPVAVTPGPAPPQHOPSPAHPPPPPVGPTTPP 592  |
| XX   | PP   | Db        | 534 VYRPPHSAPGPQPPSHAPTPPPAACAQGPQPPP -- PCPSQTTRABLPTEPAFP 590     |
| XX   | 24-MAY-1991;   | Qy        | 593 AASLPQPEAPGAEGALVNASSAAAHNVDTARAADLFVSQMGSR 637                 |
| PR   | 91US-00705814.   | Db        | 591 AATGSGPEASNAEAGALVNASSAAAHNVDTARAADLFVSQMGAR 635                |
| XX   | 07-FEB-1992;   |           |   |
| XX   | 92US-00832855.   |           |   |
| XX   | PA   |           |   |
| XX   | (ARCH) ARCH DEV CORP.  |           |   |
| XX   | PI   |           |   |
| XX   | Roizman B, Liu F;  |           |   |
| XX   | WPI: 1992-391444/48.   |           |   |
| XX   | DR   |           |   |
| XX   | New herpes serine proteases and correxp. nucleic acid sequences - for      | RESULT 14 |   |
| PT   | detection, prevention and treatment of infection caused by HSV, Epstein-   | AAR28640  |   |
| PT   | Barr, Varicella-Zoster and CMV cytomegalo-virus.                           | ID        | AAR28640 standard; protein; 626 AA.                                 |
| XX   | PS   |           |   |
| XX   | Disclosure; Table 1; 66pp; English.  | XX        |   |
| XX   | The sequences given in AAR28635-52 are mutants of the protease encoded by  | AC        | AAR28640;   |
| CC   | the UL26 open reading frame (ORF). The UL26 ORF is contained within a      | XX        |   |
| CC   | section of the herpes simplex virus (HSV) type 1 genome for the family of  | AC        |   |
| CC   | HSV capsid proteins designated ICP35. The UL26 transcription initiation    | XX        |   |
| CC   | site is used as the start point for measurements of distance in this       | DT        | 25-MAR-2003 (revised)   |
| CC   | section of the genome. The UL26 protease and ICP35 precursor proteins,     | DT        | 24-MAR-1993 (first entry)   |
| CC   | ICP35 c and d, may be cleaved by the UL26 protease. Cleavage of ICP35 c    | XX        |   |
| CC   | and d gives ICP35 e and f respectively. Cleavage specificity of the UL26   | DE        | UL26 protease deletion mutant EE, amino acids 1-9 deleted.          |
| CC   | protease is altered by the inclusion of these mutations. These mutations   | XX        |   |
| CC   | include insertions, deletions or substitutions which affect the functional | KW        | UL26; open reading frame; ORF; protease; herpes simplex virus; HSV; |
| CC   | domains of the protease. (Updated on 25-MAR-2003 to correct PN field.)     | KW        | capsid protein; ICP35; ICP35 c; ICP35 d; ICP35 e; ICP35 f;          |
| XX   | Sequence 635 AA;   | XX        | functional domain; insertion; deletion; substitution.               |
| XX   | Query Match Score 68.6%; Score 2321; DB 2; Length 635;                     | XX        | Herpes simplex virus type 1 genome for the family of                |
| Best Local Similarity 73.6%; Pred. No. 8.1e-141; Mismatches 30; Mismatches 122; Indels 18; Gaps 7; | Matches 475; Conservative  | XX        | UL26 transcription initiation                                       |
| Db   | 1 MADAQGDRMEELPDPDAVPIVAGFIALYDQGDPGELAIDPDVTVAALPPENPLPINVD 60            | XX        | CC  |
| Db   | 61 HRARCEYGRVLAVNDPRGPFYGLIACVQLERVLETAASAAIFERGPAISREERLLY 120            | XX        | CC  |
| Db   | 61 HRAGCEYGRVLAVDPRGPFYGLIACVQLERVLETAASAAIFERGPAISREERLLY 120             | XX        | CC  |
| Qy   | 121 ITNLYPSVSLSTGRGDDEPDTLFAHVALCAIGRLGTVYDTSLDAATAPFHLD 180               | XX        | CC  |
| Db   | 121 ITNLYPSVSLATKRLGGAEHPDTLFAHVALCAIGRLGTVYDTSLDAATAPFHLD 180             | XX        | CC  |
| Qy   | 121 ATREGVREAAEELAALAGRITWAQGVBAHTLTHLSTAVNNNMRLDRWSLVAERRAGIA 240         | XX        | CC  |
| Db   | 181 ASREGARLAAEELAISGRITWAQGVBAHTLTHLSTAVNNNMRLDRWSLVAERRAGIA 240          | XX        | CC  |
| Db   | 181 GHTYLOQASEKEFTWGAESAPAPERGYTGAGMAMTSSPAASVPAQ -- VAVRAQVAS 296         | XX        | CC  |
| Qy   | 241 GHTYLOQASEKEFTWGAESAPAPERGYTGAGMAMTSSPAASVPAQ -- VAVRAQVAS 296         | XX        | CC  |
| Db   | 241 GHTYLOQASEKEFTWGAESAPAPERGYTGAGMAMTSSPAASVPAQ -- VAVRAQVAS 296         | XX        | CC  |
| Qy   | 297 SSSSSPPADMNPPVSAASGAPAPPPGDSVLIWIPASHYHNLVTGOSAHPPLTACGL 356           | XX        | CC  |
| Db   | 301 S-----PVLPNNPVPTSGTAPAPRQDGSVLIWIPASHYHNLVTGOSAHPPLTACGL 354           | XX        | CC  |
| Qy   | 301 S-----PVLPNNPVPTSGTAPAPRQDGSVLIWIPASHYHNLVTGOSAHPPLTACGL 356           | SQ        | Sequence 626 AA;  |



|    |     |  |  |     |
|----|-----|--|--|-----|
| Qy | 535 | VTLPPPHIAPPGPPLS -                           | GAVPPPSYPPVAVTPGPAPPLHQPSPAHAPPPPPGPTPPP | 592 |
| Db | 534 | VRPPPHSAVGPQGPSPHAPTPPYAPACPPGPPP            | -- PCPSTQTRAPLPTBAFPP                    | 590 |
| Qy | 593 | AASLPQPSAPGAGALVNASSAAHVNVDTARAADLFVSYOMMGSR | 637                                      |     |
| Db | 591 | AATGSPDASNAEAGALVNASSAAHVNVDTARAALFVSYOMMGSR | 635                                      |     |

Search completed: January 10, 2005, 19:14:56  
Job time: 164 secs